EXPRESSION OF PROTEINS FROM CHIMERIC *TETRAHYMENA* INTRONS INTEGRATED IN RIBOSOMAL RNA GENES OF *SACCHAROMYCES CEREVISIAE*

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Robert Suran

February 2010

EXPRESSION OF PROTEINS FROM CHIMERIC *TETRAHYMENA* INTRONS INTEGRATED IN RIBOSOMAL RNA GENE OF *SACCHAROMYCES CEREVISIAE*

Robert Suran, Ph. D.

Cornell University 2010

The group I introns from *Tetrahymena thermophila*, Tth.L1925, and *Physarum polycephalum*, Ppo.L1925, are closely related, sharing 70% sequence identity in their ribozyme regions. However, Ppo.L1925 also contains an open reading frame (ORF) encoding the homing endonuclease I-PpoI, which is separated from the ribozyme by a 53 nucleotide 3'UTR region; Tth.L1925 lacks both the ORF and 3'UTR regions.

Previous studies of Ppo.L1925 have shown that the I-PpoI protein is translated from the RNA Pol I-derived spliced intron RNA. In order to better understand Ppo.L1925-encoded ORF translation, I created several chimeric ORF-containing Tth.L1925 introns and integrated them into the rDNA of *Saccharomyces cerevisiae*. Previous attempts to exchange the I-PpoI ORF found that small (~300 bp) ORFs can be integrated whereas larger ORFs disrupt intron splicing. I found that Tth.L1925-based introns could accommodate and express both small and larger ORFs; these introns produce protein at levels similar to proteins produced from wild-type Ppo.L1925 (~ 0.01% of total yeast protein).

To better understand the effect of intron sequence on protein expression, I added Ppo.L1925 sequences to chimeric Tth.L1925 introns. Some 3'UTR sequences had previously been shown to increase expression from Ppo.L1925 nearly 20-fold; however, all of the 3'UTR sequences tested in chimeric Tth.L1925 introns lowered

expression. Replacing the Tth.L1925 5'end sequence with that of Ppo.L1925, though, increased expression nearly 10-fold. To further study the effect of the intron 5' end sequence on expression, I created an intron pool with random 5' end sequences. About 25% of these introns had increased expression up to 20-fold, while about 50% had their expression lowered to undetectable levels. I examined two models for expression of chimeric intron-encoded ORFs that explain how 5' end sequence changes could affect translation: interaction with the invasive growth IRES system and translation through formation of intron RNA circles. Addition of an invasive growth IRES sequence did not increase expression from a Tth.L1925-based intron, while RT-PCR amplification of intron RNA circle junctions demonstrated no relationship between the circle species formed and expression level. These results suggest that the intron-encoded ORF is not being translated through either of these two models.

BIOGRAPHICAL SKETCH

Robert Suran was born in Gettysburg, Pennsylvania on April 24, 1974. He attended Penn State University from 1992 to 1996, where he majored in Molecular and Cell Biology. After graduating from Penn State, he came to Cornell University and joined the lab of Volker Vogt, where he studied the always interesting, and often confusing, mobile nuclear group I introns from *Physarum* and *Tetrahymena*. He took a leave of absence from graduate school in 2001 and became the workshop coordinator for the Cornell Instituite for Biology Teachers (CIBT). He worked to improve K-12 biology education at CIBT for six years. He returned to the lab finish his Ph.D. work in the fall of 2007. During his time at CIBT, he met the lovely Jantra Ngosuwan, whom he married in the summer of 2008. He has relocated to Philadelphia, PA, and is pursuing a career in science education.

For my family.

ACKNOWLEDGMENTS

I would like to thank my advisor, Volker Vogt, for his guidance and support. Volker was always positive, even when the results were confounding. I couldn't have asked for a better advisor.

I would like to thank the members of my committee, Tim Huffaker and John Lis, for providing helpful comments and insight after my student seminars and during committee meetings. I would like to thank Bik Tye for serving as a special member on my A-exam.

I am grateful to Jue Lin for all the assistance she has given me on this project and for all of our conversations about the mysteries of introns.

I thank Nicholas Zagorski for helping me edit and proofread my thesis, and also for helping me get rid of those last few pesky words in my abstract.

I would like to thank the past and present members of the Vogt lab, Danso Ako-Adjei, Andrew Babera, Becky Bean, Jany Chan, Amanda Dalton, Wayne Decatur, Rob Dick, Kari Dilley, Nicole Doria-Rose, Erin Espelie, Sharon Fodor, Marc Johnson, Swati Joshi, Paul Keller, Jue Lin, Di Lu, May Ma, Judith Phillips, Heather Scobie, and Gisela Shatz for making the lab such a wonderful home.

I would like to thank my father, John Suran, mother, Sue Feheringer, and brother, Ed Suran, for their encouragement through out the years.

Lastly, I would like to thank my wife, Jantra Ngosuwan Suran, for all her help and patience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

The Biology of Mobile Nuclear Group I Introns

Group I introns

The discovery of group I introns radically changed the role that RNA was thought to play in a cell. Until the early 1980's, RNA was thought to be an intermediary between DNA and protein. mRNA is transcribed from DNA in the nucleus, transported to the cytoplasm, where the ribosome, which was thought to be composed of catalytic protein and structural rRNA, and aminoacyl-tRNAs translate the message into protein. This view was changed in 1982 by the discovery of an RNA that could catalyze chemical reactions. Cech and his coworkers discovered that an intron in the immature ribosomal RNA (pre-rRNA) of *Tetrahymena thermophila*, currently known as Tth.L1925, could self-splice without protein factors (Kruger, Grabowski et al. 1982). This discovery was followed in 1993 by the Altman group's discovery that the RNA component of Ribonuclease P (RNase P) was responsible for RNA cleavage (Guerrier-Takada, Gardiner et al. 1983). These discoveries earned Cech and Altman the 1989 Nobel Prize in chemistry and started the search for other catalytic RNAs. Discovered ribozymes include group I introns, a family of *Tetrahymena*-like self-splicing introns. Other ribozymes include group II introns, a family of self-splicing introns unrelated to group I introns, and the RNA-cleaving hammerhead and hairpin ribozymes. *In vitro* selection of ribozymes has evolved ribozymes capable of catalyzing many reactions, such as aminoacylation of tRNA, RNA-dependent RNA polymerization, and acyl-transferase of amino-acids (Green and Szostak 1992; Lohse and Szostak 1996; Lee, Bessho et al. 2000). Proteins themselves are even the product of RNA catalysis, as the RNA in the ribosome has been shown to

be responsible for catalyzing peptide bond synthesis (Khaitovich, Mankin et al. 1999; Khaitovich, Tenson et al. 1999; Cech 2000; Nissen, Hansen et al. 2000).

Since their discovery of the *Tetrahymena* intron, group I introns have grown into a family with over 2,700 identified members (listed at http://www.rna.ccbb.utexas.edu/) (Cannone, Subramanian et al. 2002). This family is defined by a limited amount of shared primary sequence and by the ability of the intron RNA to fold into a common secondary structural motif. It is this structure that allows the intron RNA to catalyze its own removal from the surrounding RNA, preventing the intron sequence from disrupting the function of the host sequence (Cech 1988; Michel and Westhof 1990).

Group I introns share many features with Group II introns, a separate family of catalytic RNAs, and inteins, protein domains that splice themselves out of their host protein after translation. All three of these groups share similar evolutionary niches and are found interrupting transcribed DNA. They are able to post-transcriptionally (group I and group II introns) or post-translationally (inteins) catalyze their own removal from their host sequence, which allows the host gene to function despite the presence of the parasitic element. All three sequences have undergone horizontal transfer between relatively unrelated organisms. Lastly, there are many examples of these elements gaining additional sequences in the form of open reading frames (ORFs) (group I and group II introns) or separate protein domains (inteins) during the course of their evolution. These ORFs and protein domains either aid in the splicing of the element or allow it to horizontally transfer more efficiently (Perler, Olsen et al. 1997). Because of the similarities, it is illuminating to discuss group II introns and inteins alongside group I introns.

The most striking ability of group I introns, group II introns and inteins is their ability to self-splice. Each of these groups uses a distinct pathway to ensure that their

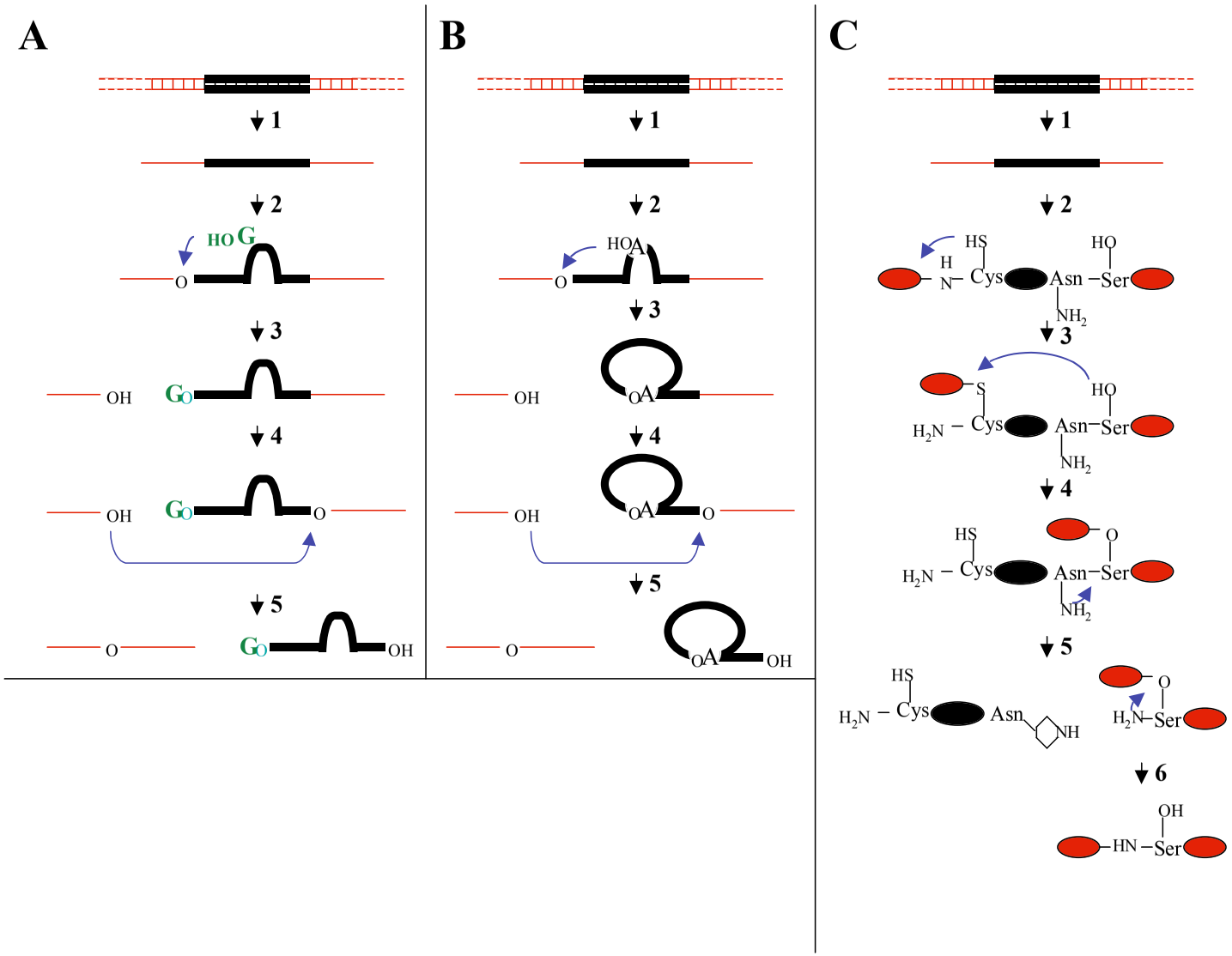
host gene functions normally. The splicing reactions are catalyzed by the insertional elements themselves, but sometimes require additional cofactors, such as an exogenous guanosine (Group I introns), Mg^{2+}, splicing factors and maturases (see intron-encoded proteins, below).

Group I Intron Splicing

In vitro, the self-splicing reaction of the many group I intron requires only Mg2+ and an exogenous guanosine cofactor (Termed the exoG, which can be a GTP, GDP, GMP, or unphosphorylated guanosine). The splicing reaction initiates with the non-covalent interaction of the exoG and the G-binding site in the P7 stem (Figure 1.1A). The 3'-OH of the exoG attacks at the 5' splice site, binding to the 5' end of the intron RNA and displacing the 3' end of the 5' exon RNA. The 3' -OH of the 5' exon RNA then attacks between the 3' end of the intron RNA and the 5' end of the 3' exon, which liberates the intron RNA and joins the exons together.

In vivo the rate of the splicing reaction is dependent on several factors. Neighboring exon RNA sequence can slow the rate of intron splicing by interfering with the folding of the ribozyme (Woodson and Cech 1991). In other cases, neighboring intron sequences can aid in folding, increasing the rate of splicing (Rocheleau and Woodson 1995; Koduvayur and Woodson 2004). Protein factors also aid group I intron slicing, increasing the rate of splicing by facilitating the folding of the ribozyme RNA (see Maturases and Splicing Factors, below).

Figure 1.1: Group I intron, group II intron, and intein splicing.
A. The group I intron splicing reaction, with the exons in red and the intron in black. 1. The intron-containing gene (DNA) is transcribed to RNA. 2. An exogenous GTP, GDP, GMP or guanosine (green) binds to the intron. 3. The 3' OH of the G residue attacks at the 5' splice site, liberating the 5' exon RNA and covalently binding to the 5' end of the intron. 4. The 3' OH of the 5' exon attacks at the 5' end of the 3' exon. 5. This attack liberates the intron and joins the 5' and 3' exons.
B. The group II intron splicing reaction with the exons in red and the intron in black. 1. The intron-containing gene (DNA) is transcribed to RNA. 2. The 2' OH of an internal A attacks at the 3' end of the 5' splice site. 3. This liberates the 5' exon and forms a RNA lariat at the 5' end of the intron. 4. The 3' OH of the 5' exon attacks at the 5' end of the 3' exon. 5. This attack liberates the intron and joins the 5' and 3' exons.
C. The intein splicing reaction. The exteins are in red and the intein in black. 1. The intein-containing gene (DNA) is transcribed to RNA. 2. The intein-containing RNA is translated to protein. 3. If the initial intein residue is a cysteine, the N-extein is transferred to the side chain of the first residue at the N-terminal end of the intein through an N-S acyl rearrangement. If the initial residue of the intein is a serine then the N-extein is transferred through an N-O rearrangement. 4. The N-terminal extein is transferred to the side chain of the C-terminal residue of the intein, which can be either serine, threonine, or cysteine, through a trans-esterification reaction. 5. The side chain of the C-terminal intein asparagine cyclize with the intein C-terminus, releasing the exteins. 6. A peptide bond is formed between the N-terminal and C-terminal exteins through a spontaneous O-N (or S-N) acyl rearrangement.


A
B
C
1
2
3
4
5
6
HO G
OH
O
HO A
OA
HS
H
N
Cys
Asn
Ser
HO
NH2
H2N
NH

Group II intron and intein splicing

The splicing of group II introns also involves two transesterification reactions (Figure 1.1B). Unlike group I introns, which use an exogenous cofactor to initiate splicing, the group II introns use the 2' OH of an internal A residue as the first attacking group of the splicing reaction. The 2'-OH of the A residue attacks at the 5' splice site, binding to the 5' end of the intron RNA and forming a "lariat" structure. This attack also displaces the 3' end of the 5' exon RNA. The 3' -OH of the 5' exon RNA then attacks between the 3' end of the intron RNA and the 5' end of the 3' exon, which liberates the intron RNA lariat and joins the exons together.

Initial observations of the splicing mechanism of group II introns and spliceosomal introns found that they were similar, leading to the suggestion that group II introns may be the forbearers of the spliceosomal complex (Sharp 1985; Cech 1986; Sharp 1991). Further mechanistic investigations of the catalysis of both of the splicing reactions have found them to be indistinguishable (Moore and Sharp 1993; Padgett, Podar et al. 1994). Furthermore, regions of group II intron sequence can complement the loss of spliceosomal RNAs and vice-versa (Hetzer, Wurzer et al. 1997; Shukla and Padgett 2002). Taken together, these facts are highly suggestive that the spliceosomal complex is another example of a protein-RNA complex taking over the role previously played purely by RNA.

Inteins, being made of protein, splice in a completely different manner, and are unrelated to both group I and group II introns (Figure 1.1C). They are translated into a protein domain that splices itself out of the host protein, or extein. The initial step of intein splicing is either an N-S acyl rearrangement if the initial intein residue is a cysteine or a N-O rearrangement if the initial residue is a serine. This results in the movement of the N-extein to the side chain of the first residue at the N-terminal end of the intein. A transesterification reaction with the first residue of the C-terminal extein,

which can be either serine, threonine, or cysteine, moves the N-extein to the side chain of the first residue in the C-extein. Cyclization of the intein C-terminal asparagine side chain excises the intein from the exteins. Lastly, a spontaneous O-N (or S-N) acyl rearrangement forms a peptide bond between the exteins (Liu 2000).

Group I intron nomenclature

There are two nomenclature schemes describing group I introns inserted in ribosomal RNA genes. The original nomenclature scheme names introns by taking the first two letters based on the organism of origin, one for genus and two for species, and follows them by three letters abbreviating the gene within which they are inserted and a number indicating the order in which the intron was discovered. For example, the group I intron from the large ribosomal subunit of *Physarum polycephalum* would be named PpLSU3: "Pp" for the first letters of ***P****hysarum* ***p****olycephalum*, "LSU" for the **l**arge ribosomal **s**ub**u**nit, and "3" as it was the third intron found in this gene and organism (Dujon, Belfort et al. 1989; Michel and Westhof 1990).

More recently, nomenclature has been proposed for introns inserted in genes encoding ribosomal RNA (rDNA). The newer nomenclature takes into account the position of insertion based on the numbering of the *E. coli* sequence. The *Physarum* intron would be named Ppo.L1925: "Ppo" for ***P****hysarum* ***po****lycephalum*, "L" for **l**arge ribosomal subunit, and "1925" as it is inserted between 1925 and 1926 in reference to the *E. coli* LSU sequence. An additional letter is added before the subunit letter to denote location in mitochondrial (an "m") or chloroplast (a "c") genome, as in the mitochondrial intron Sce.**m**L2449 (Johansen and Haugen 2001). Unfortunately, this nomenclature has not been widely accepted. In order to avoid confusion, this thesis uses the common names of each intron and the above nomenclature for the *Tetrahymena* intron (Tth.L1925) and *Physarum* intron (Ppo.L1925).

Group I intron structure

Group I introns are able to fold into a very similar secondary structure even though there is often less than ten percent sequence identity between many of them (Davies, Waring et al. 1982; Michel, Jacquier et al. 1982; Cech, Damberger et al. 1994). This structure has been confirmed though chemical and enzymatic probing (Been and Cech 1987; Burke 1988; Couture, Ellington et al. 1990) and later through crystal structures of introns from *Tetrahymena*, *Azoarcus*, and bacteriophage *Twort*. (Golden, Gooding et al. 1998; Adams, Stahley et al. 2004; Adams, Stahley et al. 2004; Guo, Gooding et al. 2004; Golden, Kim et al. 2005).

The secondary structure of a group I ribozyme (Figure 1.2A) consists of regions of paired RNA sequence, denoted by "P" followed by a number indicating the order of the stem, 5' to 3'. The core ribozyme region, which is defined as being both the conserved sequences among group I introns and the sequences necessary for ribozyme activity, consists of P1, P3, P4, P5, P6, P7, P8, and P9 regions. Other regions, such as P5abc (Figure 1.2B), P2-P2.1 and P9.1-P9.2 in the case of the Tth.L1925 pictured, are not present in all cases (Michel and Westhof 1990). Based on the presence or lack of these optional regions, group I introns have been classified into five major classes, A through E. These are divided into smaller subgroups based on variations within the group (Michel and Westhof 1990).

Crystal structures show that the core region of the ribozyme is arranged into two long helical domains consisting of the P5-P4-P6 and P8-P3-P7-P9 stem loops. The P1 region, which contains of the 3' end of the 5' exon and the 5' splice site, forms a third helix that interacts with the interface of the two larger stem loops. The intron sequence that pairs with this region, known as the internal guide sequence (IGS), is responsible for determining the location of the 5' end of the intron. Changes to the IGS can either change the location of the 5' splice site or abrogate splicing entirely.

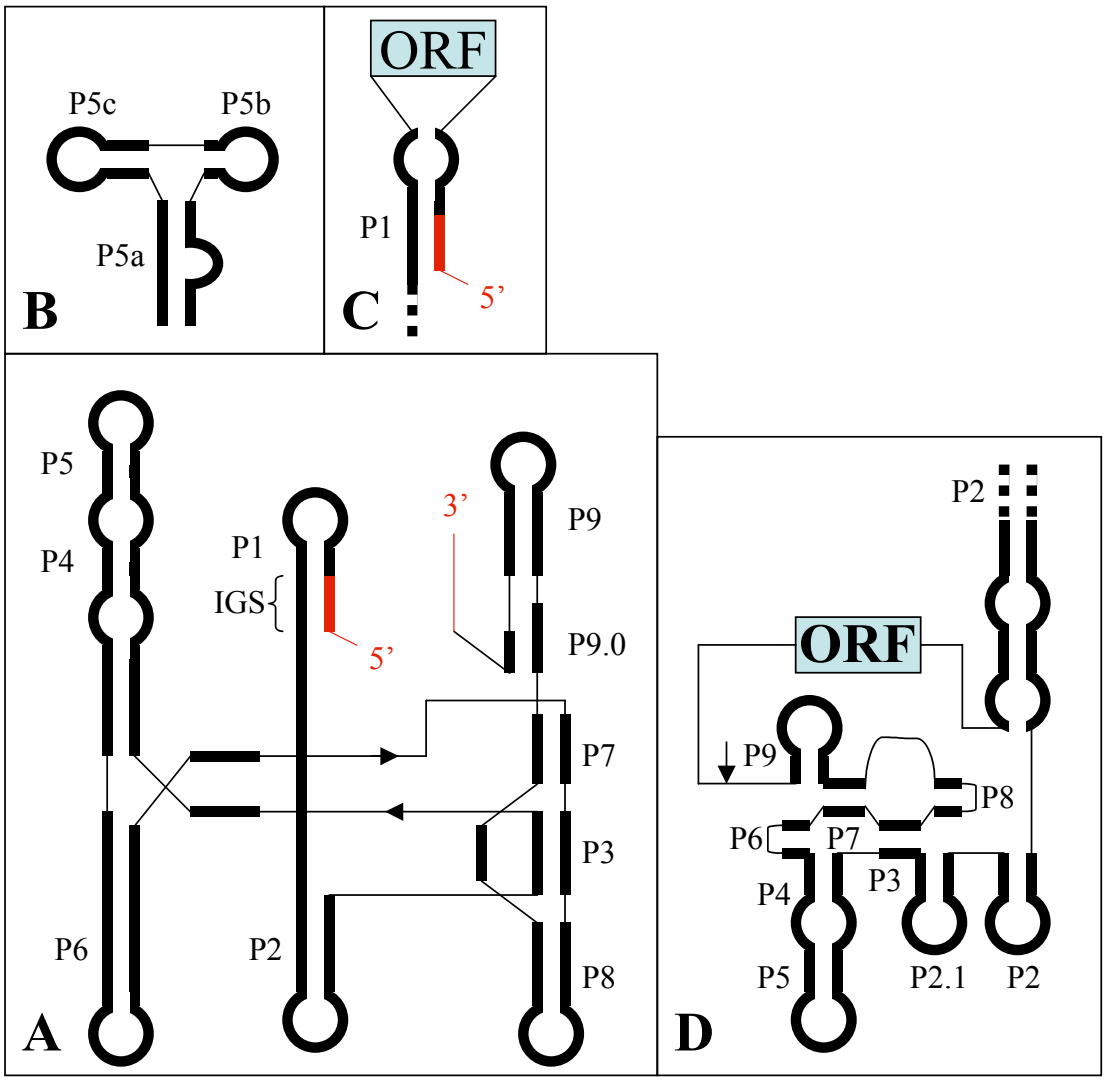


Figure 1.2: Group I intron structure.
A. The core structure of a group I intron. The intron is in black, while the exon sequences are in red.
B. The extend P5abc loop found in some group I introns. The P5abc loop stabilizes the three-dimensional structure of the intron.
C. Some group I introns contain ORF sequences in loops. The group I intron Ppo.L1925 contains an ORF in the P1 loop as shown here.
D. Some group I introns have an ORF and a second group I ribozyme inserted into the P2 loop.

The G binding site is located in the P7 stem, while the P9 and P10 regions play a role in determining the location of the 3' splice site (Golden, Gooding et al. 1998).

Mg^{2+} is necessary for group I intron activity *in vitro*, both by playing roles in stabilizing the RNA structure and in the chemistry of the splicing reaction (Pyle 1993; Hanna and Doudna 2000). Three Mg^{2+} ions are proposed to play a role in the ribozyme catalysis, two of which coordinate the nucleophile in both steps of the splicing reaction, while the third stabilizes the leaving group (Weinstein, Jones et al. 1997). Candidates for the former two Mg^{2+} ions are observed in the *Azoarcus* intron structure. These two ions interact with all three helical domains and bring the 5' and 3' splice sites into the catalytic core (Adams, Stahley et al. 2004).

Group I introns often contain peripheral sequences in addition to the core ribozyme. These sequences include RNA that folds to support the intron structure, ORFs that are inserted within loop regions of the introns, or even a sequence that forms a second ribozyme structure. Many introns contain extended RNA sequence in the P5 loop, the P5abc structure (Figure 1.2B), which helps to stabilize the intron structure (see Maturases, below). Inserted ORFs have been found in both sense and anti-sense orientation and usually code for maturases or homing endonucleases, proteins that give the intron an advantage either by aiding in intron splicing or in the transfer of the intron (Figure 1.2C). Two introns, Nja.S516, and Dir.S956-1, have a twin-ribozyme organization (Figure 1.2D) (Johansen and Vogt 1994; Decatur, Einvik et al. 1995). Cleavage by the second ribozyme plays a role in expression of the intron-encoded ORF (See Expression of Nuclear Group I Intron Encoded Genes, below).

Post-splicing reactions

Group I intron RNAs also undergo a number of post-splicing reactions, some of which seem to have additional biological significance. One of these reactions is a reverse of the splicing reaction, where the intron splices into another RNA (Figure

1.3A). Initiation of reverse splicing is primarily mediated by IGS binding. The intron RNA can splice into both the exon RNA and other RNA molecules with sequence complementary to the IGS. As group I intron splicing is essentially a two-step chemical reaction, consisting of two transesterifications catalyzed by a ribozyme with no net energy change, the splicing reaction is fully reversible. Spliced full-length intron will cleave a target RNA and insert itself into RNA at sites homologous the site of intron insertion. The new insertion site does not necessarily have to be the original site of intron insertion; it merely has to have homology to the original site for IGS binding. Weaker IGS binding results in a slower reaction. A sequence neighboring the insertion site can also interfere with reverse splicing if it forms secondary structures that block IGS binding. Reverse splicing has been demonstrated both *in vitro* (Woodson and Cech 1989) and *in vivo* (Thompson and Herrin 1994; Roman and Woodson 1995; Birgisdottir and Johansen 2005) and may play a role in the horizontal transfer of group I introns lacking homing endonucleases (see Horizontal Transfer of Group I Introns).

Some group I intron ribozymes catalyze cleavage of the spliced intron RNA (Figure 1.3B). These processing cleavages are mediated by a G addition in a similar manner to the first step of splicing. The 3-OH of the exoG attacks at the cleavage point, cleaving the intron into a 5' piece and a 3' piece with an additional 5' G. This processing has been shown to reduce protein production from Ppo.L1925 in yeast and may serve to reduce the levels of potentially toxic homing endonuclease produced from the intron (Ruoff, Johansen et al. 1992; Lin and Vogt 1998; Decatur, Johansen et al. 2000; Haugen, De Jonckheere et al. 2002).

Group I introns also form circular RNA after splicing (Figure 1.3D and 1.3E). The circularization reaction can be thought of as a reverse of the last step of the splicing reaction, wherein the 3' -OH of the intron attacks the 5' end of the intron.

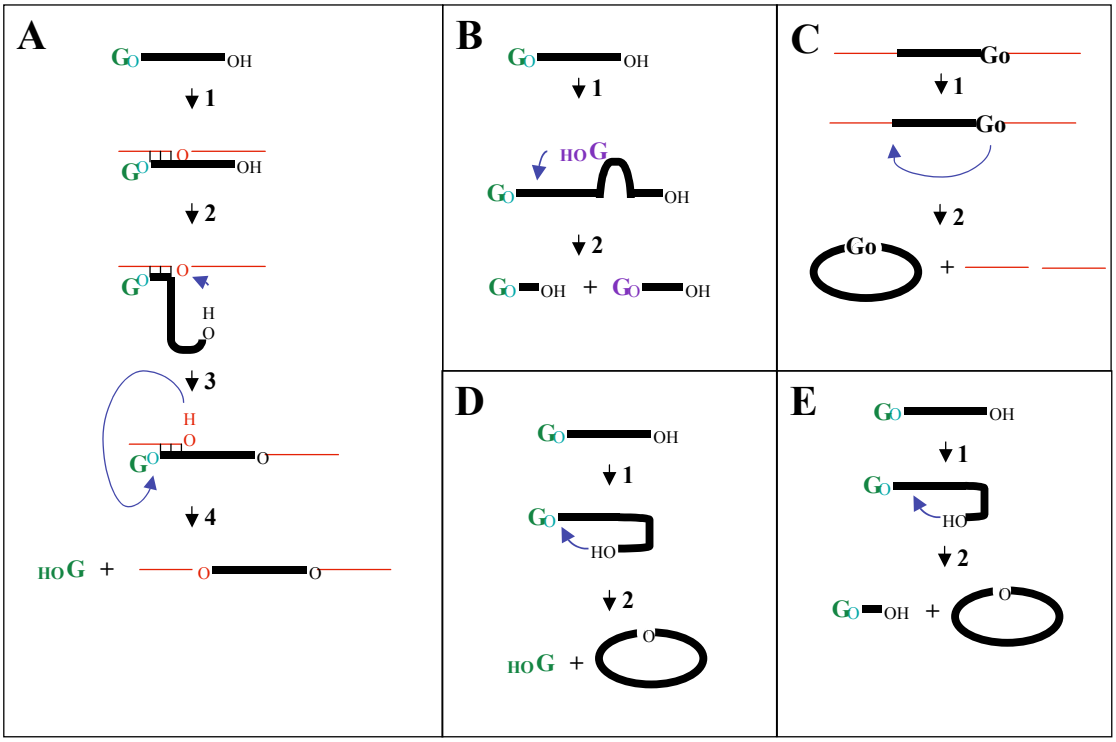


Figure 1.3: **Post-splicing reactions**.
A. Reverse splicing: 1. The IGS of the fully-spliced intron (black) binds to a free target RNA (red). 2. The 3' OH of the intron attacks the bound RNA downstream of the IGS binding site. 3. The 3' OH of the 5' target RNA attacks between the exogenous G (green) added in the splicing reaction and the first nucleotide of the intron. 4. This frees the exogenous G and inserts the intron into the target RNA.
B. Processing. 1. A second exogenous G (purple) binds to the spliced intron RNA (black). The 3' OH of this G attacks in the intron sequence. 2. This results in the cleavage of the intron and the second exogenous G covalently bound to the 5' end of the 3' intron piece.
C. Formation of full-length intron circles pre-splicing. 1. The terminal G of the intron attacks at the 5' splice site. 2. This results in the formation of a full-length intron circle and does not ligate the 5' and 3' exons (red lines).
D. Formation of full-length intron circles post-splicing. 1. The 3' OH of at the 3' end off the spliced intron (black) attacks between the exogenous G (green) added in the splicing reaction and the first nucleotide of the intron. 2. This frees the exogenous G and circularizes the intron RNA.
E. Formation of truncated intron circles. The 3' OH of at the 3' end off the spliced intron (black) attacks in the intron sequence. 2. This frees the 5' end of the intron sequence and circularizes the 3' intron RNA.

There appear to be two general types of circles formed, full-length circles where the 3' end of the intron attacks between the exogenous G and the first nucleotide of the intron, and truncated circles, where the 3' end of the intron attacks in or near the IGS of the intron. Tth.L1925 primarily forms IGS circles at the +15 and +19 positions both *in vitro* and *in vivo*, although full-length circles have been seen as a minor species *in vitro* (Been and Cech 1986; Inoue, Sullivan et al. 1986; Been and Cech 1987).

The group I intron Dir.S956-1 forms full-length circles through an alternate pathway, which occurs before the intron splices itself out of its host RNA (Figure 1.3C). The bond between the 3'-most nucleotide of the intron, a G (referred to as the ΩG), and the 5' end of the 3' exon is hydrolyzed. The 3' OH of the ΩG attacks at the 5' splice site, resulting in a circular intron. Evidence for the formation of full-length pathways has been seen for most nuclear group I introns *in vitro*, including Tth.L1925. However, this pathway seems to be inhibited in Tth.L1925 *in vivo*. Interestingly, this pathway is detrimental to the intron's host, as circle formation by this pathway produces nonfunctional rRNA (Nielsen, Fiskaa et al. 2003).

Intron encoded and associated proteins

The majority of intron-encoded proteins serve one of two functions. They either are homing endonucleases, which allow the intron to enter into intron-lacking alleles, or are maturases, proteins that aid in the intron splicing reaction. There are exceptions, though, such as the *Neurospora* mitochondrial large rRNA intron, which encodes the essential ribosomal S-5 protein (Burke and RajBhandary 1982).

Homing endonuclease genes (HEG) have been found in group I introns, group II introns, inteins, and in inter-genic regions (Belfort and Roberts 1997). HEGs allow their parent insertional element to transfer, or home (as DNA), into element-lacking alleles through a double-strand break repair (DSBR) mechanism (see Horizontal transfer of group I introns, below). The endonucleases these genes produce are similar

to restriction enzymes in that they recognize a specific DNA target site and make a double stranded break at that site, although homing endonuclease recognition sites are much larger, ranging from12 to 40 bp. Homing endonucleases do not necessarily recognize the strict palindromic sequences that are characteristic of restriction enzyme target sites and they tolerate sequence degeneracy in their target site, which is uncommon in restriction enzymes. Intron-encoded homing endonucleases also tend to be smaller than restriction enzymes, and it is possible that their size is constrained by requirements for intron splicing. There are four families of homing endonucleases based on active site sequence motifs: His-Cys box, LAGLIDADG, H-N-H, and GIY-YIG.(Lambowitz and Belfort 1993; Mueller 1993; Roberts 1993; Roberts and Macelis 1997).

Homing endonucleases are named in a fashion similar to restriction enzymes with the addition of a prefix to indicate the location of the endonuclease ORF. Intron-encoded endonucleases are indicated with an “I” prefix, intein-encoded endonucleases with a "PI", and freestanding ORF's are indicated by an "F" prefix. The endonuclease encoded by the *Physarum* intron Ppo.L1925 is named I-*Ppo* I, “I” as it is intron-encoded, “Ppo” for ***P****hysarum* ***po****lycephalum*, and “I” as it is the first discovered homing endonuclease encoded by an intron in this organism (Roberts, Belfort et al. 2003).

The reaction rate of self-splicing introns *in vitro* is often much slower than the intron splices *in vivo*. *In vivo*, proteins can bind to the intron and increase the intron’s reaction rate. When these proteins are intron-encoded, they are called maturases, while proteins encoded elsewhere in the genome are termed splicing factors.

Group I intron maturases appear to have evolved from LAGLIDADG motif homing endonucleases (Belfort and Roberts 1997). Many group I intron encoded proteins possess both maturase and endonuclease activity, while a few amino acid

substitutions can activate endonuclease activity in the I-*Cre* I maturase (Dalgaard, Klar et al. 1997). Crystal structures of the I-*Ani*I, a bi-functional homing endonuclease/maturase from the *Aspergillus nidulans* mitochondrial COB intron (AnCOB), show that the maturase and endonuclease activity are the result of distinct domains. It is possible that maturase activity developed in regions of the protein not important in endonuclease function after the HEG sequence entered into the intron sequence (Lambowitz 1989; Bolduc, Spiegel et al. 2003; Chatterjee, Brady et al. 2003).

Splicing factors are host-encoded proteins that aid in intron splicing. There are two well-characterized group I intron splicing factors, *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase Cyt-18 and the yeast cytochrome B processing protein Cbp2. Other known splicing factors include two proteins from yeast, Mrs2 and leucyl-tRNA synthetase NAM2 and two *E. coli* proteins, StpA and ribosomal protein S12.

Cyt-18 is a mitochondrial tyrosyl-tRNA synthetase in *Neurospora crassa*. It can also act as a splicing factor for three mitochondrial group I introns, the mitochondrial large ribosomal RNA intron, cytrochrome oxidase intron *cob*-I2, and NADH dehyrdogenase intron ND1 (Akins and Lambowitz 1987; Wallweber, Mohr et al. 1997). CYT-18 has also been shown to suppress splicing mutations in the T4 td intron core that abrogate splicing (Mohr, Zhang et al. 1992) and can facilitate the splicing of the yeast bI5 intron (Webb, Rose et al. 2001). All of the previously mentioned introns lack an extension of the P5 stem, the P5abc region which is present in other group I introns, such as the *Tetrahymena* intron, Tth.L1925. The P5abc region enhances intron folding, and therefore enhances splicing, by stabilizing the P4-P6 stem. A crystal structure of CYT-18 bound to the Twort ribozyme shows that the protein serves in a similar capacity in P5abc-lacking introns by binding to the P4-P6

region of the ribozyme (Paukstelis, Chen et al. 2008). CYT-18 does not stimulate the splicing of P5abc-containing Tth.L1925 *in vitro* but does rescue splicing in a mutant Tth.L1925 lacking the P5abc region (Mohr, Caprara et al. 1994; Wallweber, Mohr et al. 1997).

The yeast CBP2 is a nuclear gene that plays the role of splicing factor with several yeast introns. Unlike CYT-18, CBP2 does not appear to have a function in addition to that of a splicing factor. CBP2 was first isolated in a screen for respiratory mutants in *Saccharomyces cerevisiae*, and was later found to act as a splicing factor for the last intron in the mitochondrial cytorchrome oxidase b, *b*I5 (McGraw and Tzagoloff 1983; Hill, McGraw et al. 1985; Gampel, Nishikimi et al. 1989). The lack of phenotype in yeast that lack both CBP2 and *b*I5 led to a search for another function for the protein. *Saccharomyces douglasii* was found to encode a CBP2-like protein, but lacks the *b*I5 intron, suggesting that CBP2 had an additional activity, which resulted in its conservation in *S. douglasii*. Investigation showed, though, that CBP2 was acting as a splicing factor in *S. douglasii*, with it being required for the splicing of a group I intron in the mitochondrial large ribosomal subunit gene (Tian, Li et al. 1998). CBP2 stimulates the splicing of a similar *S. cerevisiae* intron, Sce.mL2449 *in vitro*, but is not essential for splicing of this intron *in vitro* or *in vivo*. Unlike CYT-18, CBP2 does not accelerate splicing in either the AnCOB intron or the td intron (Shaw and Lewin 1997).

Studies using *b*I5 intron as a model show that CBP2 facilitates splicing of introns in a different fashion than CYT-18. These proteins interact with different areas of the intron; CYT-18 interacts primarily with the P5-P4-P6 domain, while CBP2 interacts with the P1-P2 and P3-P7.1-P8 regions of the intron (Hill, McGraw et al. 1985; Weeks and Cech 1995; Caprara, Lehnert et al. 1996; Buchmueller, Webb et al. 2000; Webb, Rose et al. 2001). CYT-18 binds when the intron RNA is in a

collapsed but unfolded state. The initial state does not involve specific RNA-CYT-18 or RNA-RNA interactions. The unstructured ribozyme starts to fold, initially at the P5-P4-P6 domain, with which CYT-18 interacts. The remaining RNA regions then slowly fold around the P5-P4-P6 region to form the active CYT-18-bI5 complex (Webb, Rose et al. 2001). CBP2, on the other hand, does not interact the bI5 RNA until after the RNA has moved from the unstructured collapsed state to a state where the bI5 core is nearly formed, (Weeks and Cech 1996; Buchmueller, Webb et al. 2000). After CBP2 binding, the intron RNA folds rapidly, and in a single step, to an active state (Webb and Weeks 2001).

Mrs1 and Nam2 act as splicing factors for the third and fourth cytochrome oxidase introns, *b*I3 and *b*I4 respectively, in yeast. These proteins act in conjunction with intron-encoded maturases to facilitate splicing. In the case of the *b*I3 intron, both the splicing factor and maturase appear to be proteins that have been co-opted by the intron to aid in splicing and have lost their original functions. Mrs1 is a nuclear protein which shares homology with the Holiday Junction protein Cce1, but is inactive as a nuclease, while the *b*I3-encoded maturase contains a the LAGLIDADG motif, but is inactive as an endonuclease (Kreike, Schulze et al. 1986; Kreike, Schulze et al. 1987; Dalgaard, Klar et al. 1997; Wardleworth, Kvaratskhelia et al. 2000). Mrs1 binds the *b*I3 intron RNA as a tetramer, along with a single copy of *b*I3 maturase, to form a large RNS-protein complex that is competent for splicing (Bassi, de Oliveira et al. 2002).

Nam2 is a yeast mitochondrial leucyl-tRNA synthetase that is involved in splicing two yeast mitochondrial introns, *b*I4 and cytrochome oxidase a intron, *a*I4 (Labouesse, Herbert et al. 1987; Herbert, Labouesse et al. 1988). Both of these introns encode maturases, but mutations in the *b*I4 maturase that disrupt its maturase activity block the splicing of both *b*I4 and *a*I4 introns. Mutations in Nam2 can restore splicing

activity to both introns. Interestingly, splicing of these introns in the mutant Nam2 strains is now dependent on *a*I4 maturase, as disruption of the *a*I4 maturase ORF disrupts the splicing activity of both introns. Although the mechanism that the *a*I4 maturase uses to compensate for the loss of *b*I4 is currently unclear, three-hybrid analysis of Nam2, *b*I4 maturase and *b*I4 intron has shed light on how the three interact with each other. This analysis indicated that Nam2 and the maturase are not directly interacting with each other. Instead, the intron RNA bridges between the two proteins (Rho and Martinis 2000).

Two *E. coli* proteins have been found to act as splicing factors for the T4 phage *td* intron. StpA, a transcriptional repressor, was isolated because it rescued a splicing defective *td* intron and stimulates *td* splicing *in vitro* (Zhang, Derbyshire et al. 1995). *E. coli* ribosomal protein S12 has also been shown to facilitate splicing of *td* intron *in vitro* and *in vivo* (Coetzee, Herschlag et al. 1994; Semrad and Schroeder 1998). Unlike the previously mentioned splicing factors, which recognize and bind specific RNA structures in the intron, StepA and S12 both bind RNA non-specifically. These proteins act as RNA chaperones and prevent misfolding of the intron RNA or aid in melting misfolded RNAs (Herschlag 1995; Clodi, Semrad et al. 1999).

Horizontal transfer of group I introns

Group I introns are found in a variety of genes in bacteria, chloroplasts, mitochondria, viruses, and the nuclei of fungi and algae (Lambowitz and Belfort 1993). Bacterial, viral, chloroplast, and mitochondrial group I introns occupy positions within genes that encode proteins, rRNA or tRNA. Nuclear group I introns, though, are invariably found within genes encoding ribosomal RNA. The reason for this localization is not well understood, but suggests that group I introns depend on one or more of the special properties of rDNA, either during the horizontal transfer event, or later, to aid in intron splicing. These nucleolus-specific properties include

the many protein and RNA factors that localize to the nucleolus, the high sequence conservation of rDNA between organisms (the sites of intron insertion tend to be in the most conserved regions. See (Johansen, Haugen et al. 2007) for a summary of known nuclear group I introns insertion sites.), or the arrangement of rRNA genes in a large number of tandem repeats (usually over 100 copies).

Group I introns occupy these diverse positions because they are mobile elements, able to move to intron-lacking loci in different species. Evolutionary trees based on intron sequence vary widely from those constructed using the surrounding exon sequence. For example, a comparison of the intron and rDNA sequences of several species of *Tetrahymena* found that the intron Tth.L1925 was independently acquired on multiple occasions (Nielsen and Engberg 1985; Sogin, Ingold et al. 1986). In addition, Tth.L1925 shares a very strong sequence similarity with other group I introns. It shares 70% identity with its closest known relative, the ribozyme of Ppo.L1925 (Ruoff, Johansen et al. 1992). This implies that group I introns have transferred horizontally between relatively different organisms. Similar results have been found with introns in fungi (Holst-Jensen, Vaage et al. 1999), red and brown algal ribosomal RNA (Bhattacarya, Cannone et al. 2001) and introns in the *cox1* gene of plant mitochondria (Cho, Qiu et al. 1998).

Mobility is commonly conferred to group I introns through intron-encoded endonucleases. These endonucleases allow the intron to enter into intron-lacking alleles through a process called homing (Figure 1.4). Homing endonucleases recognize a relatively large sequence (~12-40 nucleotides) around the site of intron insertion in intron-lacking alleles. They catalyze a double-stranded break at the site of intron insertion. Repair of the break through the DSBR pathway, using the intron-containing allele as a template, results in the intron being copied into both alleles. Due to the use of the DSBR pathway, intron homing also results in a variable extent of co-

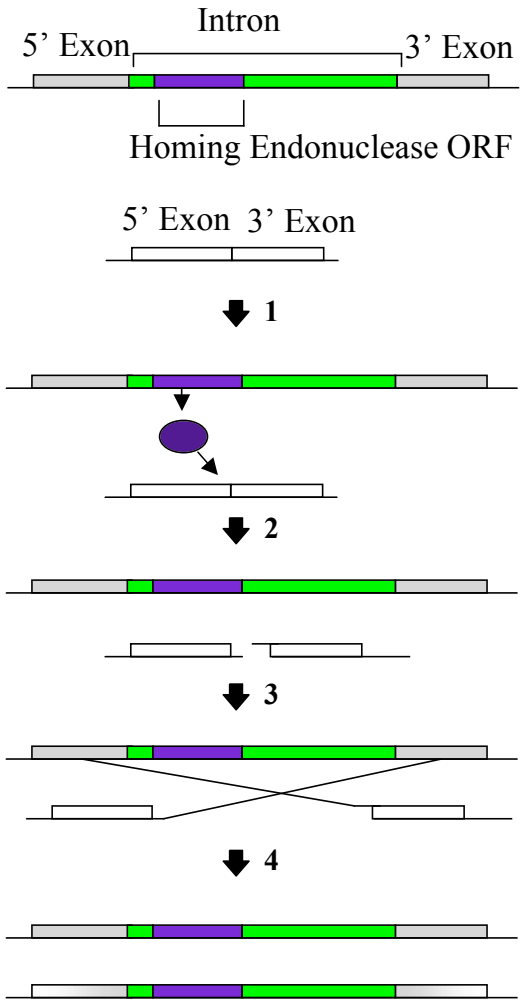


Figure 1.4: Intron homing. 1. Homing endonuclease (purple circle) is produced from the intron-containing allele. 2. The homing endonuclease makes a double-stranded break in the intron-lacking allele. 3. The intron-containing allele is used as a template for repair of the intron-lacking allele through double-stranded break repair. 4. This results into the intron being copied into the intron lacking allele, and some amount of co-conversion of the 5’ and 3’ exon sequence.

conversion of flanking sequences around the site of intron insertion. The insertion of the intron interrupts the homing endonuclease target site, making the repaired allele resistant to further cleavage (Belfort and Perlman 1995).

Homing was first studied in Sce.mL2449, also known as Ω, an intron in yeast mitochondrial LSU RNA . To date, homing is the only demonstrated mechanism of horizontal transfer of group I introns (Colleaux, d'Auriol et al. 1986; Colleaux, D'Auriol et al. 1988). Ppo.L1925 was the first nuclear group I intron to be demonstrated to be mobile. Ppo.L1925 homing has been demonstrated both in strains of *Physarum*, where it is found naturally, and in *S. cerevisiae*, where it can be induced artifically. In yeast, the intron is able to home both from a plasmid encoded-sequence into the chromosomal XII rDNA loci and from intron-containing chromosomal loci to intron lacking chromosomal loci (Muscarella and Vogt 1989; Muscarella and Vogt 1993).

The fact that the majority of group I introns lack encoded homing endonucleases raises the question of how these introns entered into their respective alleles. There are two possibilities to explain their distribution. Either these introns once possessed homing endonucleases that were later lost, or they were able to enter into their current position through a different mechanism. One model is transfer through a retrotransposition mechanism. In this model, the intron reverse splices into a new site in an RNA, followed by the conversion of the RNA to DNA by an endogenous reverse transcriptase, and reinsertion of the DNA into a chromosome by homologous recombination (Belfort and Perlman 1995; Holst-Jensen, Vaage et al. 1999; Birgisdottir and Johansen 2005).

The transposition model of horizontal transfer has not been demonstrated at the present, although it is similar to the retrohoming and retrotransposition of group II introns, wherein an intron-encoded protein plays the role of reverse transcriptase (see

below) (Eskes, Yang et al. 1997; Cousineau, Smith et al. 1998; Eskes, Liu et al. 2000). Evidence pointing to group I intron transposition includes partial and full reverse splicing into new target sites *in vivo* (Woodson and Cech 1989; Thompson and Herrin 1994; Roman and Woodson 1995; Birgisdottir and Johansen 2005). No group I introns encode their own reverse transcriptase, so it must come from another source, possibly a retrovirus or retrotransposon. It is also unclear how the reverse transcribed RNA would make its way into the nucleus to undergo recombination. Nevertheless, the idea of group I introns horizontally transferring through reverse transcriptase-mediated process is appealing, as not only does it explain how the introns can transfer from the same gene between species, but it also how they can enter into new sites.

The larger question is the identity of the vector that allows transfer between organisms. It is difficult to imagine how an intron is transferred from the chloroplast genome of one angiosperm to one in a distantly related angiosperm. At the moment, there is no evidence to explain this transfer, although it has been suggested that transfer could occur through feeding of one organism on another or through the intermediate of a pathogenic fungus or a double stranded RNA virus (Holst-Jensen, Vaage et al. 1999).

Group II Intron Retrohoming and Retrotransposition.

Unlike ORF-encoding group I introns, group II introns enter into new sites through a RNA-mediated process called retrohoming (Figure 1.5). Although there are several variant strategies for group II intron retrohoming, all of them utilize intron-encoded multifunctional proteins that have maturase, endonuclease and reverse transcriptase activity. For example, horizontal transfer of the yeast *COXI* intron, aI2, is initiated by the binding of the intron-encoded protein to the spliced intron lariat to form a ribonucleoprotein (RNP) complex. The RNP complex makes a nick in the intron-lacking DNA by intron-catalyzed partial reverse splicing of the intron RNA

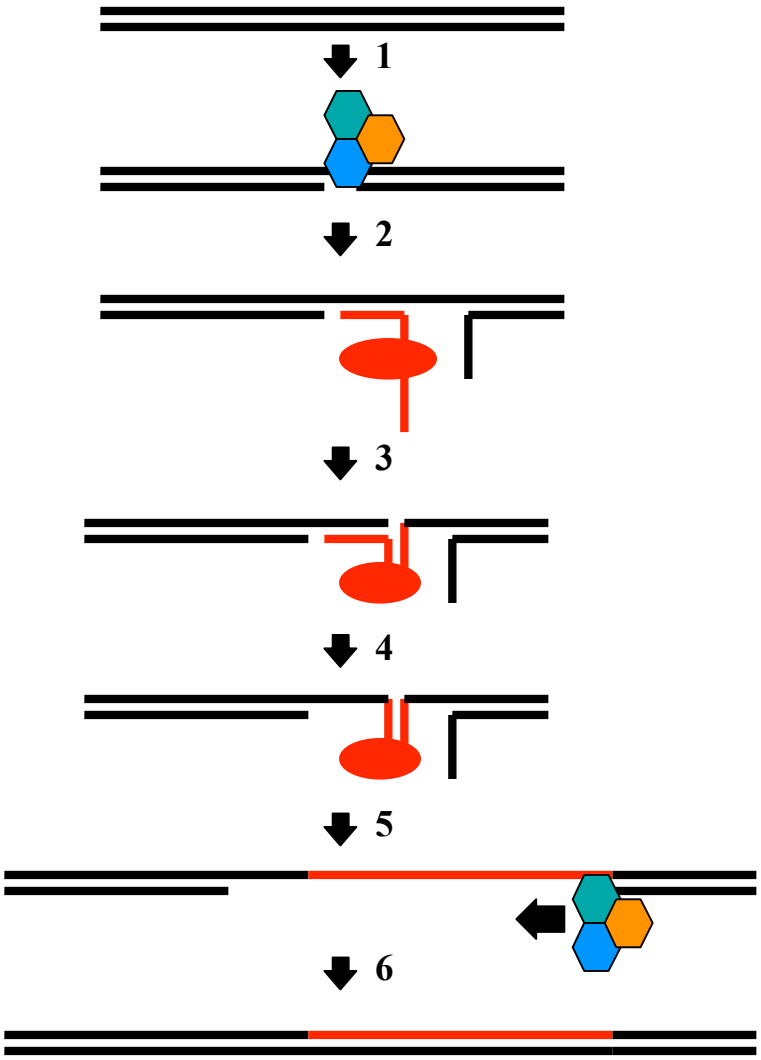


Figure 1.5: Reverse splicing. 1. The intron-encoded protein (three circles, contains reverse transcriptase (RT-green), maturase (M-orange) and endonuclease (E-blue) activities) causes a single stranded break in the antisense strand of DNA (black lines). 2-4. The intron RNA (red) binds to the IGS at the cut site and through reverse splicing, inserts itself into the sense strand. 5. The intron RNA is then converted to DNA in the antisense strand by the RT activity of the intron encoded protein. 6. This results in the permanent integration of the intron into the new locus.

onto the 5' end of the DNA that codes for the 3' exon. The second strand of the target DNA is then cleaved by the exonuclease domain of the RNP. The reverse transcriptase domain of the RNP fills in the break, priming at the 5' end of the second nick and using an unspliced pre-mRNA as a template. Use of the pre-mRNA template results in co-conversion of sequence 5' of the intron, but minimal to no co-conversion of sequence 3' of the intron (Moran, Zimmerly et al. 1995; Zimmerly, Guo et al. 1995; Zimmerly, Guo et al. 1995)**.**

The retrohoming of the yeast *COXI* intron aI1 follows the initial steps as the neighboring aI2 intron, except that the intron RNA attached to the target DNA is used as a template for reverse transcription. This RNA can be used either as a template for full repair, which results in no co-conversion of upstream or downstream sequences, or the intron RNA can be partially copied and the rest of the gap is filled in using the DNA-mediated DSBR pathway, resulting in co-conversion of 5' sequences but not 3' sequences. Interestingly, a small percentage (20%) of aI1 intron horizontal transfers appear to be RT-independent homing events similar to group I intron homing. As with group I intron homing, these events result in co-conversion of both upstream and downstream sequences (Eskes, Liu et al. 2000).

The aI1 intron has been shown to be able retrotranspose into ectopic sites in the *COX1* gene that share sequence homology with the original insertion site of the intron position. The mechanism of intron transposition is similar to retrohoming, with the RNP nicking the ectopic site. The intron RNA reverse-splices fully into the ectopic DNA site, which is then repaired by the intron-encoded RT. As only the reverse spliced intron serves as the template for the RT, there is no co-conversion of the upstream or downstream sequence at the ectopic site with sequence from the donor site (Dickson, Huang et al. 2001).

Horizontal transfer of homing endonuclease genes

Homing endonucleases confer mobility upon the elements that they are inserted in, but they also appear to be mobile elements in and of themselves. The evidence for this mobility includes the sporadic distribution of endonucleases, with similar ORFs being found in group I introns, group II introns, and inteins. Endonuclease ORFs are also often optional a particular intron, with some strains of host organism having ORF-less introns, while other strains have introns that encode endonucleases. There are several examples of homing endonucleases colonizing new sequences, some of which are discussed below (Mota and Collins 1988; Shub, Gott et al. 1988; Bell-Pedersen, Quirk et al. 1990; Haugen, Wikmark et al. 2005).

Two examples occur in introns from the fungus *Podospora*. The first involves a group I intron in the mitochondrial *nad1-I4* gene, which is found in one of three states, without an ORF, with a single endonuclease ORF, or with two endonuclease ORFs. The second ORF, a homing endonuclease, is able to move into one-ORF introns when protoplasts from two different strains were fused. Analysis of single nucleotide differences between the two introns shows that, while the sequence near the site of the second ORF's insertion is converted from that of the one-ORF intron to that of the two-ORF intron, the sequence farther away remains that of the one-ORF intron. Thus, the transfer only involves the ORF and does not involve replacement of the entire one-ORF intron with a two-ORF intron (Sellem and Belcour 1997).

The second *Podospora* example involves a mitochondrial group I intron from *Podospora curvicolla* which can encode two endonuclease ORFs, one ORF which is present in all copies of the intron, and a second, optional ORF. Heteroplasmon formation experiments between the two strains resulted in the second ORF sequence being transferred to the allele that lacked it. The second endonuclease cleaves at or

near the ORF insertion point in the target intron, and appears to be transferred to the single-ORF sequence through a homing process (Saguez, Lecellier et al. 2000).

Another fascinating example of HEG mobility is the case of *Didymium iridis* introns Dir.S956-1 and Dir.S956-2. Even though both of these introns are inserted at the same small ribosomal subunit gene in two closely related strains of *D. iridis*, the Dir.S956-1 and Dir.S956-2 splicing ribozymes are only very distantly related. Dir.S956-1 has a twin ribozyme structure and is a group IC ribozyme, while Dir.S956-2 has only a single ribozyme, which belongs to group IE. The I-DirI ORF is inserted between the two ribozymes in Dir.S956-1 and contains a 51 nucleotide spliceosomal intron, while the I-DirII ORF is inserted in the anti-sense direction related to Dir.S956-2, and contains a 50 nucleotide spliceosomal intron. The I-DirI and I-DirII genes themselves are very closely related. The intron sequence flanking the I-DirII ORF and the exon sequence flanking Dir.S956-2 are identical in 14 out of 17 positions. This implies that the ancestral Dir.S956-2 sequence would have been recognized by I-DirII and suggests that the I-DirII ORF was acquired through a homing mechanism (Haugen, Wikmark et al. 2005).

Expression of nuclear group I intron encoded genes

To date, all nuclear, ORF-containing group I introns found have been inserted in genes that code for ribosomal RNA, or rDNA. This poses a potential paradox, as rRNA is transcribed by RNA polymerase I (RNA Pol I), while eukaryotic proteins are translated from RNAs that are transcribed by RNA polymerase II (RNA Pol II). RNA polymerase II transcripts have distinctive features that distinguish them from the normally untranslated RNA Pol I and III transcripts (Figure 1.6). These features includes a 7-methyl G cap on the 5' end of the RNA and 3' poly-(A) tail. Capped and polyadenylated RNAs are translated much more efficiently than corresponding RNAs lacking either a cap or tail (Gallie 1991).

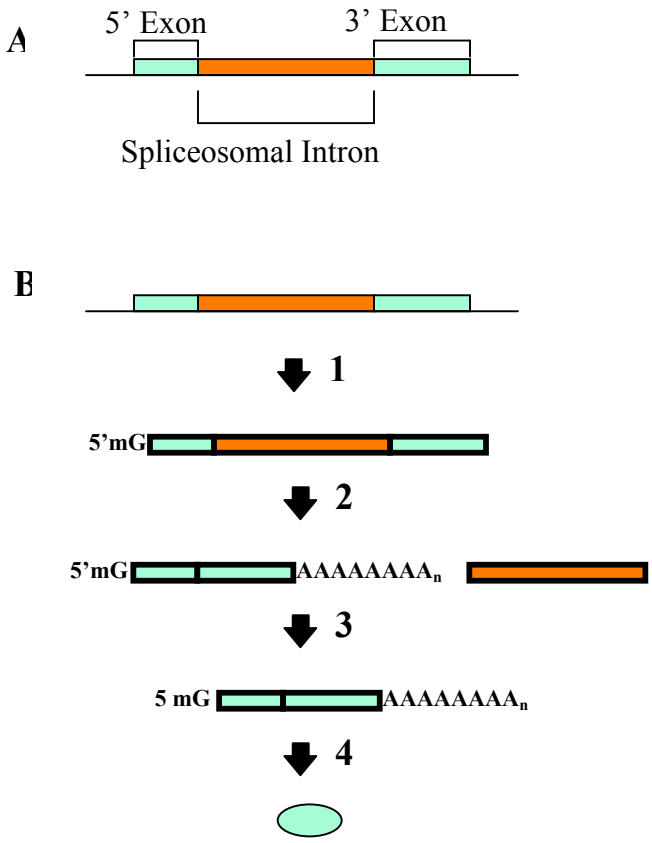


Figure 1.6: Expression of a generic RNA Polymerase II-transcribed gene. A. A typical RNA polymerase II gene. B. Expression of a typical RNA Polymerase II-transcribed gene. 1. The DNA (thin lines) is transcribed by RNA Polymerase II into RNA (thick lines). A 5'methyl-G cap is added co-transcriptionally. 2-3. The spliceosomal complex catalyzes the splicing of the intron out of the upstream and downstream RNA. A Poly-a tail is added to the 3' end of the RNA. 4. The RNA is translated into protein.

In translation initiation, the cap is specifically bound by eukaryotic initiation factor 4E (eIF4E), which interacts with eIF4G. eIF4G recruits eIF3 and the 40S ribosomal subunit (Altmann, Sonenberg et al. 1989; Morino, Hazama et al. 1996; Tarun and Sachs 1997; Tarun, Wells et al. 1997; Sachs and Varani 2000). The poly-(A) tail contributes to this interaction as well. It is bound by the poly (A)-binding protein (Pab1p) (Sachs, Davis et al. 1987). Pab1p also helps to recruit the ribosome by interacting with eIF4G (Imataka, Gradi et al. 1998; Kessler and Sachs 1998).

Endonuclease expression from six nuclear group I introns has been examined in detail, revealing several different strategies for protein production. Ppo.L1925 has been studied primarily in *S. cerevisiae*, since *S. cerevisiae* is a more tractable organism that *P. Polycephalum*. The intron can be artificially trans-integrated into the yeast genome by transforming two plasmids into the yeast, one encoding Ppo.L1925 and the other encoding the I-*Ppo* I ORF under GAL promoter control (Figure 1.7). Inductionof the homing endonuclease leads to double stranded breaks in all the rDNA copies in the yeast genome. This kills the majority of the yeast cells. The cells can survive in two different ways. Some yeast cells encode mutations in their rDNA that cause them to be resistant to cleavage by the I-PpoI endonuclease. These mutations rapidly spread to the other ~120 copies of rDNA after I-PpoI induction. Other cells survive, in a manner similar to intron homing, by using the plasmid-encoded intron as a template for repair. This results in the insertion of the intron into all of the rDNA copies. In order for the yeast cells with inserted introns to survive, the transcribed intron RNA must be functional enough to able to be splice itself out of the pre-rRNA. Yeast with integrated non-functional introns will be unable to make ribosomes and will die. It is assumed that when introns are unable to be integrated, this implies that that intron is not capable of efficiently splicing *in vivo* (Muscarella and Vogt 1993).

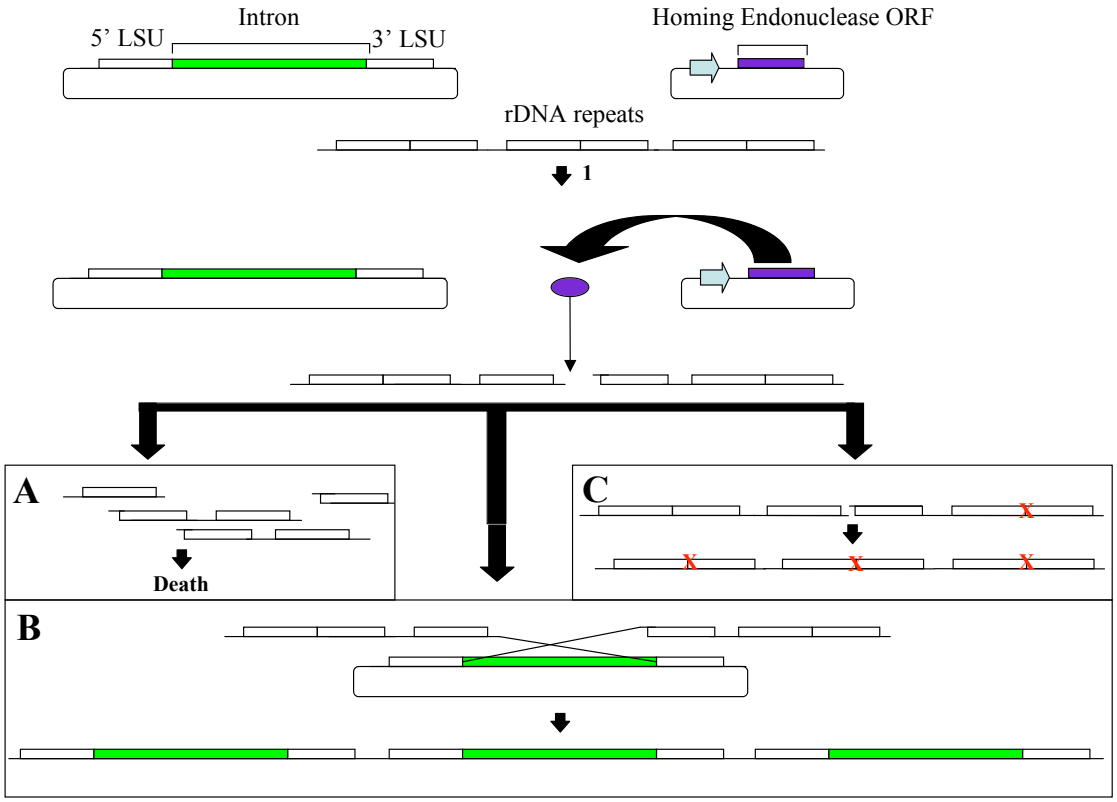


Figure 1.7: Trans-integration. A plasmid containing the intron of interest, flanked by ribosomal RNA sequence, and a plasmid with a homing endonuclease ORF under Gal promoter control are transformed into *S. cerevisiae.* 1. Production of the endonuclease is induced and it makes double stranded cuts in the rRNA on chromosome 12. The double stranded cuts are: A. Not repaired, leading to the death of the host yeast cell. B. Repaired using the intron as a template, which results in the intron being spread into all the rDNA copies in the yeast. Or C. Repaired using a rDNA copy containing an endonuclease-resistant mutation, which results in the mutation being spread into all the rDNA copies of the yeast.

The Ppo.L1925 ORF is located in the P1 stem of the intron, and is followed by a 53 nucleotide region termed the 3'UTR, which lies between the ORF and the resumption of ribozyme sequence (Figure 1.8). In *S. cerevisiae* Ppo.L1925 undergoes post splicing at two points, Internal Processing Site 1 (IPS1) which is 53 nt from the end of the I-*Ppo* I ORF and Internal Processing Site 2 (IPS2), which is 15 nt downstream from IPS1. The IPS1 cleavage occurs both in *P. Polycephalum* and in *S. cerevisiae*, and defines the end of the 3' UTR. Cleavage at IPS2 occurs at the base of the P1 stem and only occurs in *S. cerevisiae*. Additionally, Ppo.L1925 has been shown to form full length intron circles *in vitro*, although whether these are formed *in vivo* and if they are biologically significant is unknown (Haugen, De Jonckheere et al. 2002) (Figure 1.9).

While Pol I-transcribed intron RNA was an obvious source of RNA for translation of I-*Ppo* I, there was also the distinct possibility that the I-*Ppo* I message was produced from a cryptic RNA Pol II promoter. An elegant experiment showed that I-*Ppo* I is translated from a Pol I-derived RNA (Figure 1.10). Ppo.L1925 was trans-integrated into a yeast strain with a *ts* RNA Pol I. This strain does not normally grow at restrictive temperatures, due the lack of ribosomal RNA, but a plasmid encoding the 35S rRNA under *GAL7* promoter control can provide enough rRNA to compensate for the lack of Pol I. If the I-*Ppo* I message was translated from a Pol II message, the protein should still be produced when the yeast was shifted to the restrictive temperature. The strain produced no I-*Ppo* I at the restrictive temperature, and thus the I-*Ppo* I message must be transcribed by RNA Pol I or somehow otherwise be dependent of RNA Pol I translation of the rDNA locus (Lin and Vogt 1998).

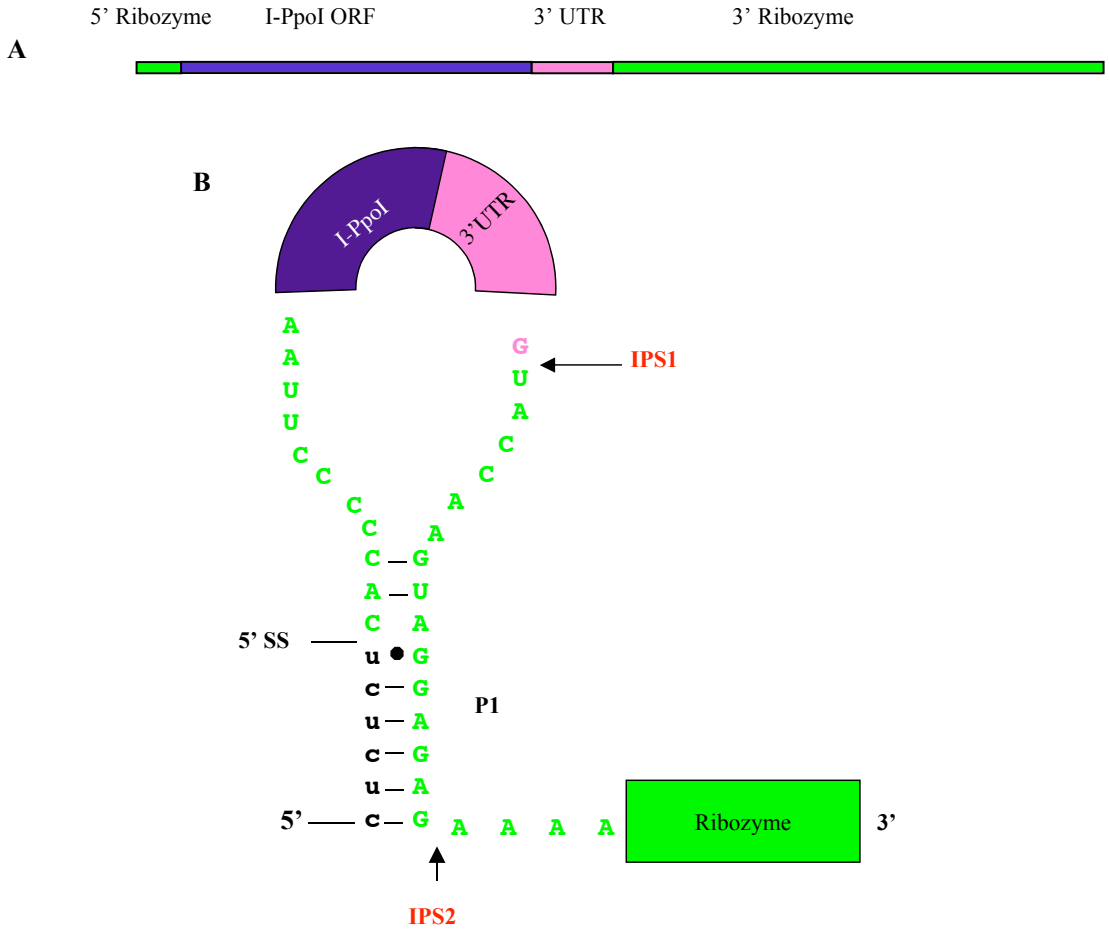


Figure 1.8: Ppo.L1925. A. Representation of important Ppo.L1925 regions. Note that there is a small region (11 nucleotides) of ribozyme RNA 5' of the I-PpoI start site. B. Diagram of the Ppo.L1925 P1 stem-loop. The P1 stem-loops consists of the 5' exon sequence (black), Ppo.L1925 ribozyme sequence (green), the I-PpoI ORF (purple box) and the 3'UTR (pink). IPS1 defines the end of the 3'UTR. IPS2 cleavage only occurs at in *S. cerevisiae*.

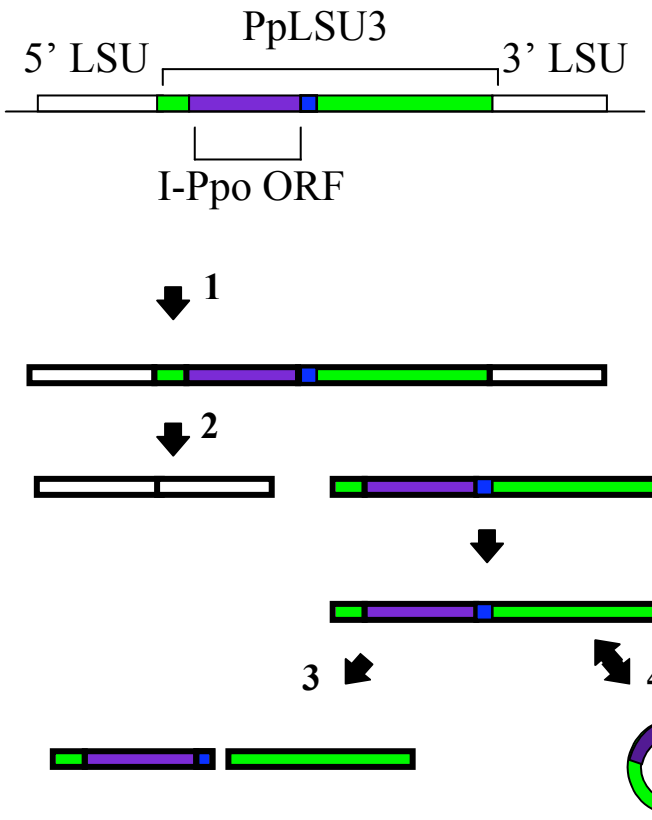
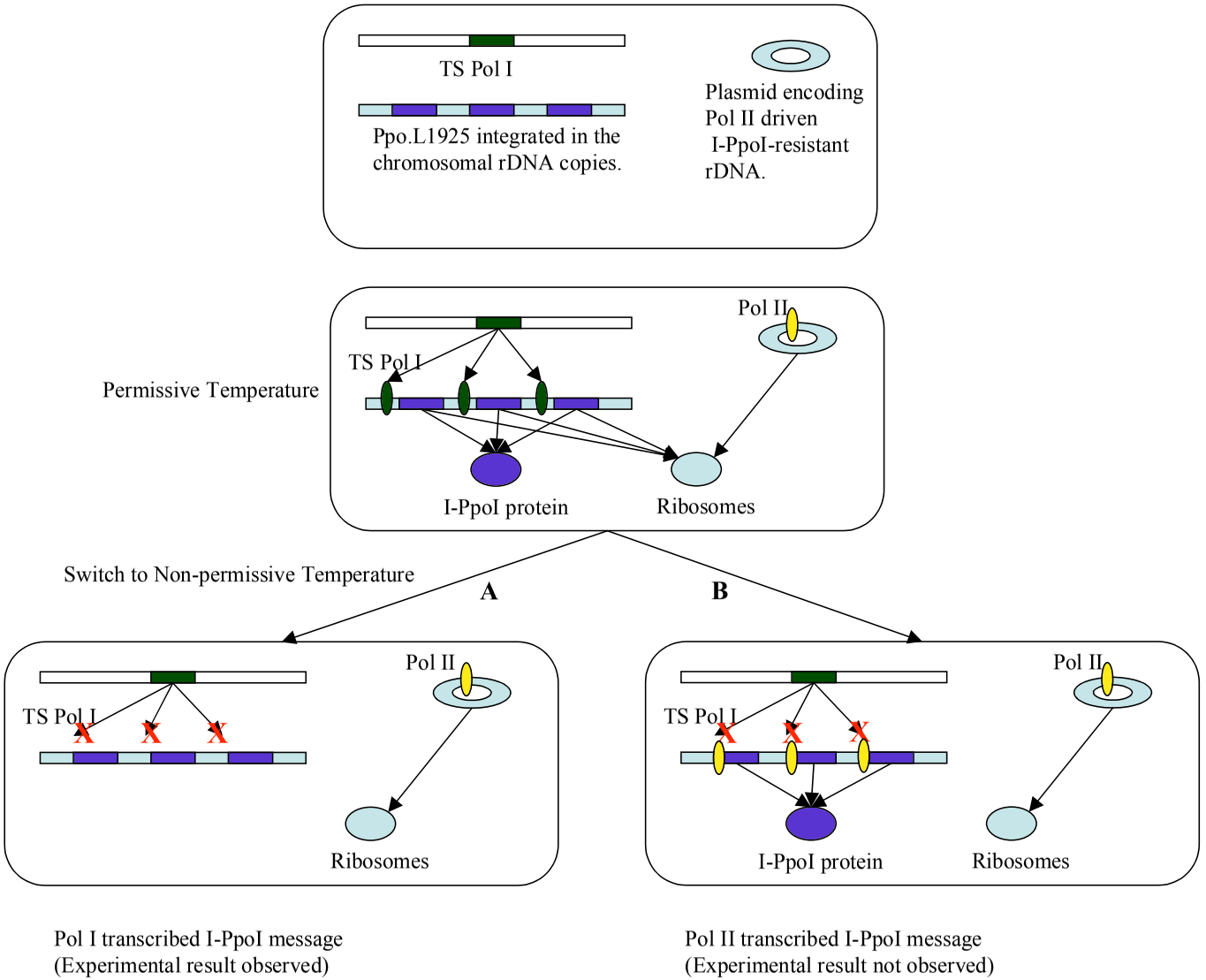


Figure 1.9: The splicing and post-splicing processing of Ppo.L1925. 1. Ppo.L1925 is transcribed by RNA Polymerase I as part of the pre-rRNA. 2. The intron splices itself out of the pre-rRNA. 3. The full-length intron is processed at either IPS1 or IPS2. 4. The intron circularizes to form either a full-length or internal circle. The circularization reaction is reversible, and full-length circles can open up to return to linear, full-length RNA.

Figure 1.10: I-PpoI is translated from a RNA Pol I-transcribed RNA. Ppo.L1925 is integrated into the chromosomal rDNA of a yeast strain with a temperature sensitive RNA Polymerase I. This yeast is transformed with a plasmid encoding ribosomal RNA under RNA Polymerase II control. When the yeast is shifted to the non-permissive temperature, the TS Pol I is not functional and the Pol II plasmid provides all of the rRNA of the yeast cell. A. If I-PpoI is translated from a Pol I-transcribed RNA, no I-PpoI should be produced at the restrictive temperature. This is the result seen. B. If I-PpoI were translated from an RNA transcribed from a cryptic Pol II promoter, I-Ppo I should still be produced at the restrictive temperature.



Pol I transcribed I-PpoI message
(Experimental result observed)

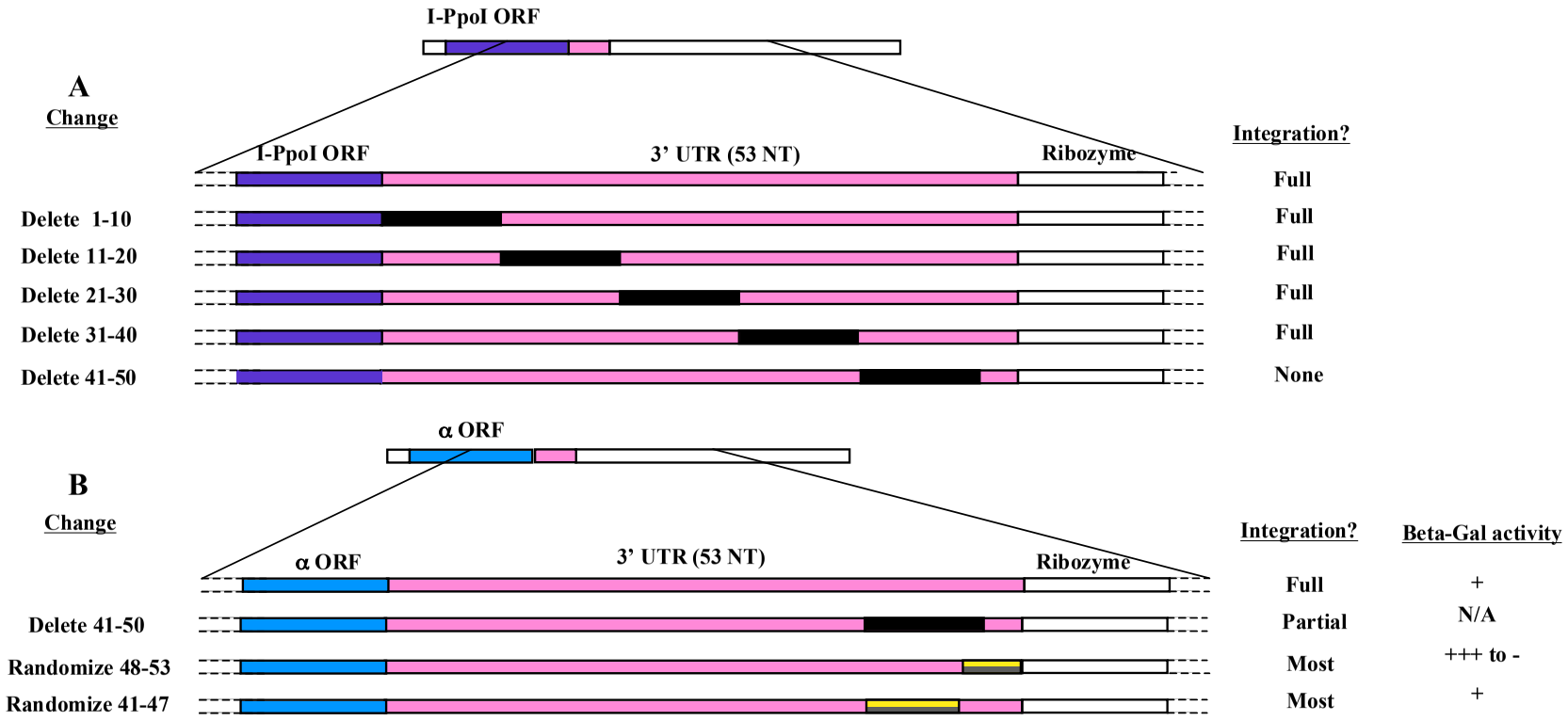
Pol II transcribed I-PpoI message
(Experimental result not observed)

The function of the Ppo.L1925 3'UTR sequence is mysterious. Initially it was thought that post-splicing cleavage at IPS 1 was important in translation of the intron-encoded ORF, as that cleavage removes the ribozyme region from the ORF. This is not the case, as mutation of IPS1, which abolishes processing and which should also reduce or prevent expression of the intron-encoded ORF if the processed RNA served as the message, actually increased translation of the intron-encoded ORF. This suggests that the full-length intron RNA actually serves as the I-*Ppo* I message. The role of the processing is unknown, but could be a way to reduce the expression of the homing endonuclease, which, at high levels, could potentially cut chromosomal DNA at degenerate sites, to levels that are more manageable for the cell (Lin and Vogt 1998).

The nucleotides at the 3' end of the 3' UTR play an important role in both splicing and translation. The importance of the nucleotides on the 3' end of the 3' UTR in splicing was demonstrated by a series of mutant Ppo.L1925 introns deleting the 3' UTR in 10 nucleotide segments (Figure 1.11A) (Lin and Vogt 2000). Deleting nucleotides 1-10, 11-20, 21-30, or 31-40 appeared to have not effect on integration of Ppo.L1925 in *S. cerevisiae*. Deleting nucleotides 41-50 prevented the mutant Ppo.L1925 from integrating, suggesting that those nucleotides are important for splicing.

The construction of Ppo-α, a chimeric Ppo.L1925, provided more data on expression from the intron and the function of the 3' UTR (Lin and Vogt 2000). The Ppo-α was constructed by replacing the I-*Ppo* I ORF with the sequence for the β-galactosidase α fragment. When the β-galactosidase α fragment forms a complex with the larger Ω fragment, it serves as a reporter protein when both are expressed *in trans*. When normalized for the relative levels of transcript, the β-galactosidase

Figure 1.11: Role of the Ppo.L1925 3'UTR in splicing and intron-encoded protein production. A. Role of 3' UTR in Ppo.L1925. Deletion of nucleotides 1-10, 11-20, 21-30, and 31-40 (black bar) does not prevent the intron from integrating into the yeast genome. Deletion of nucleotides 41-50 prevents the intron from integrating. B. Role of the 3' UTR in Ppo-α. Integration Ppo-α lacking nucleotides 41-50 results with either yeast containing both intron-integrated and I-PpoI resistant rDNA or yeast containing only I-PpoI resistant rDNA.

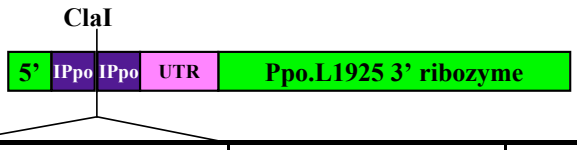

I-PpoI ORF
A
Change
I-PpoI ORF
3' UTR (53 NT)
Ribozyme
Integration?
Full
Delete 1-10
Full
Delete 11-20
Full
Delete 21-30
Full
Delete 31-40
Full
Delete 41-50
None
α ORF
B
Change
α ORF
3' UTR (53 NT)
Ribozyme
Integration?
Beta-Gal activity
Full
+
Delete 41-50
Partial
N/A
Randomize 48-53
Most
+++ to -
Randomize 41-47
Most
+

message of the intron-containing yeast was found to be translated at roughly 3% of the rate of the β-galactosidase α fragment translated from an RNA Pol II promoter.

Through the ease of assays of β-galactosidase activity, this chimeric intron allowed for larger scale testing of expression from the intron (Figure 1.11B). Deletion of nucleotides 41-50 of the 3' UTR of Ppo-α also reduces the ability of the intron to integrate into the yeast genome. The intron is able to integrate into some, but not all, of the ~120 copies of the rDNA with 3' UTR nucleotides 41-50 deleted. Presumably, the remaining copies of rDNA contain I-PpoI resistance mutations that are able to supply enough rRNA to the yeast to live.

The effect of 3' UTR nucleotides 41-53 on splicing and translation was tested through the creation of two pools of mutant Ppo-α introns. One pool, Ppo-α ran1, contained introns with 3' UTR nucleotides 48-53 randomized, while the other pool, Ppo-αran2, contained introns with randomized bases at positions 41-47 of the 3'UTR. If the primary sequence in this region is important for splicing, very few of the randomized introns should have been able to integrate into the yeast genome. This was not the case, as the majority of the randomized introns were able to integrate. Additionally, while yeast with integrated introns from Ppo-α ran2 exhibited β-galactosidase activity similar to that of Ppo-α, the majority of yeast with integrated Ppo-αran1 introns exhibited higher levels of β-galactosidase. Surprisingly, a few Ppo-α introns had over 40 fold more activity than the wild type Ppo-α. There is no obvious consensus sequence shared by these high expressor, or HE, 3' UTR sequences. Although how the 3'UTR is affecting translation is still mysterious, this data suggests that, like the IPS1 cleavage, one of the roles of the 3' UTR is to down-regulate translation of the homing endonuclease (Lin and Vogt 2000).

Attempts to make other chimeric Ppo.L1925 introns by replacing the I-PpoI ORF with other ORFs have been largely unsuccessful (Figure 1.12). There are



Intron	ORF Insert	Length of insert	Integration?
Ppo-Neo	Neo	822 bp	No
Ppo-URA3	URA 3	807 bp	No
Ppo-HIS3	HIS3	678 bp	No
Ppo-TRP1	TRP1	678 bp	No
Ppo-α	β-galactosidase α fragment	255 bp	Yes
Ppo-SOM1	SOM1	225 bp	Yes

Figure 1.12: Chimeric Ppo.L1925 introns. ORFs were inserted into Ppo-ClaI, a version of Ppo.L1925 with the majority of the I-PpoI ORF deleted and replaced with a ClaI site. Fifteen I-PpoI ORF nucleotides were retained 5' and 3' of the ClaI.

exceptions, an intron containing the 225 nt SOM1 ORF, which encodes a mitochondrial protease (Esser, Pratje et al. 1996), was able to be integrated into the yeast genome (Nam 1999). However, chimeric introns containing the *HIS3*, *URA3*, *TRP1*, or Neo ORFs were unable to integrate into the yeast genome. This is most likely due the size of the ORFs, with the larger ORFS interfering with the ribozyme-mediated splicing of the intron.

In summary, the I-PpoI message is inferred to be translated from the spliced, full-length intron RNA, while post-splicing processing reduces the available I-PpoI message (Figure 1.13).

Two introns from *Naegleria*, Nae.L1926 and Nmo.L2563, are very similar to Ppo.L1925 (Figure 1.14). They both undergo processing at sites very similar to those of Ppo.L1925 and form full-length intron circles *in vitro*. Nae.L1926 and Nmo.L2563 have 13 nt 3' UTRs, which is much smaller than 53 nt Ppo.L1925 3' UTR. Both Nae.L1926 and Nmo.L2563 3' UTRs encode polyadenylation signals which are absent in the Ppo.L1925 3' UTR. This implies that in this case processing of the intron is important for endonuclease expression (Haugen, De Jonckheere et al. 2002).

Dir.S956-1, found in the small ribosomal subunit of *Didymium iridis* strain Pan2, is a mobile group I intron with a twin-ribozyme organization (Figure 1.15) (Johansen and Vogt 1994; Decatur, Einvik et al. 1995; Johansen, Elde et al. 1997). The intron is organized with the I-*Dir*I ORF sandwiched between the two ribozymes, GIR1 and GIR2. GIR2 is responsible for catalyzing intron excision, exon ligation, and circularization of the spliced intron. GIR1 is a group I branching ribozyme that catalyzes the formation of a 5'-2' lariat at the 5' end of the I-*Dir*I ORF. This branching reaction results in a cleavage between the GIR1 sequence and the ORF. (Decatur, Einvik et al. 1995; Einvik, Decatur et al. 1997; Nielsen, Westhof et al. 2005).

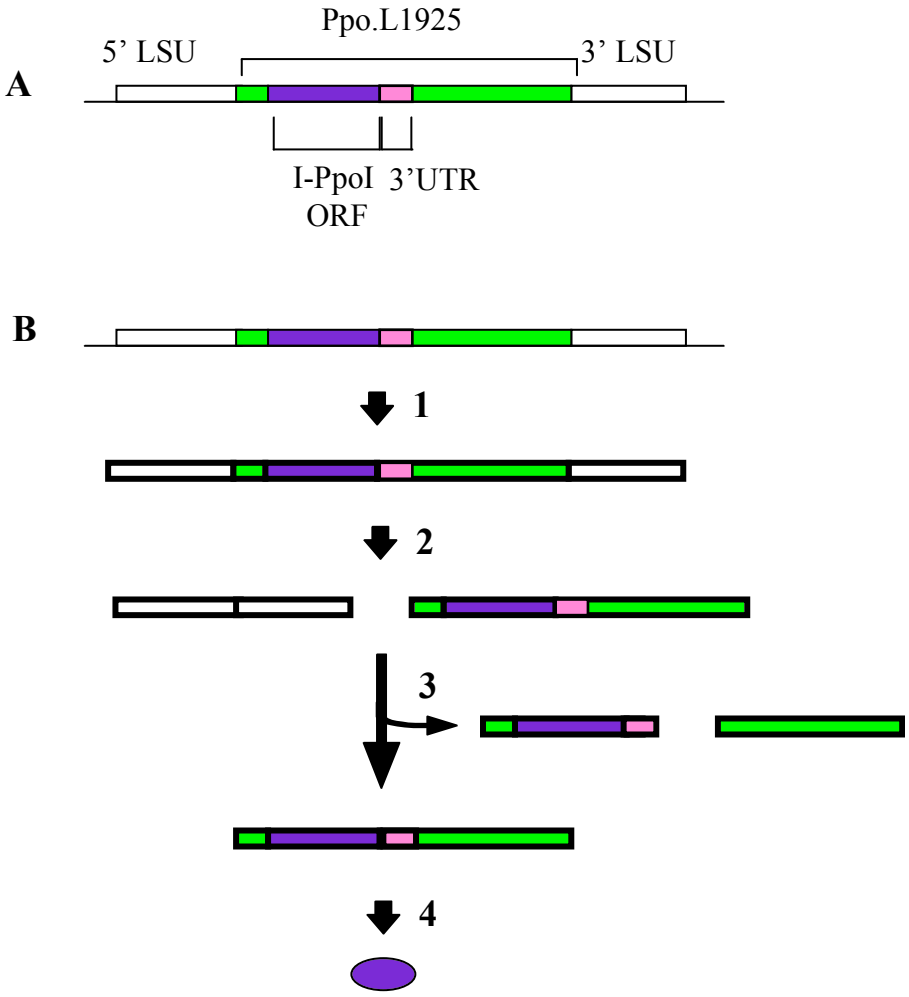


Figure 1.13 Expression of I-PpoI from Ppo.L1925. A. A schematic representation of Ppo.L1925. B. Expression of I-PpoI from Ppo.L1925. 1. The DNA is transcribed by RNA Polymerase I into RNA. 2. The Ppo.L1925 ribozyme catalyzes the splicing of the intron out of the upstream and downstream RNA. 3. Some of the introns are processed, removing them from the pool of potential I-PpoI message. 4. I-PpoI is translated from the full-length intron.

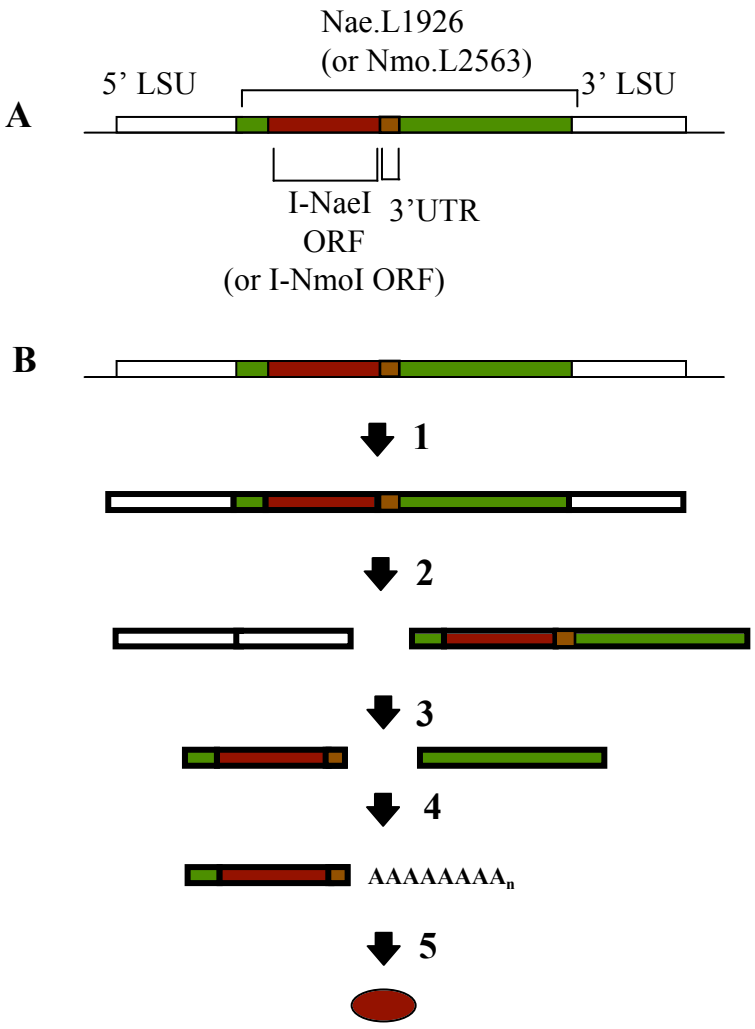


Figure 1.14 Expression of I-NaeI from Nae.L1926. A. A schematic representation of Nae.L1926 (Nmo.L2563 is similar). B. Expression of I-NaeI from Nae.L1926 (or I-NmoI from Nmo.L2563). 1. The DNA is transcribed by RNA Polymerase I into RNA. 2. The intron ribozyme catalyzes the splicing of the intron out of the upstream and downstream RNA. 3. The intron RNA molecule is processed. 4. Processing of the intron reveals a polyadenylation site, which results in the polyadenylation of the intron RNA. 5. I-NaeI is translated from the processed RNA.

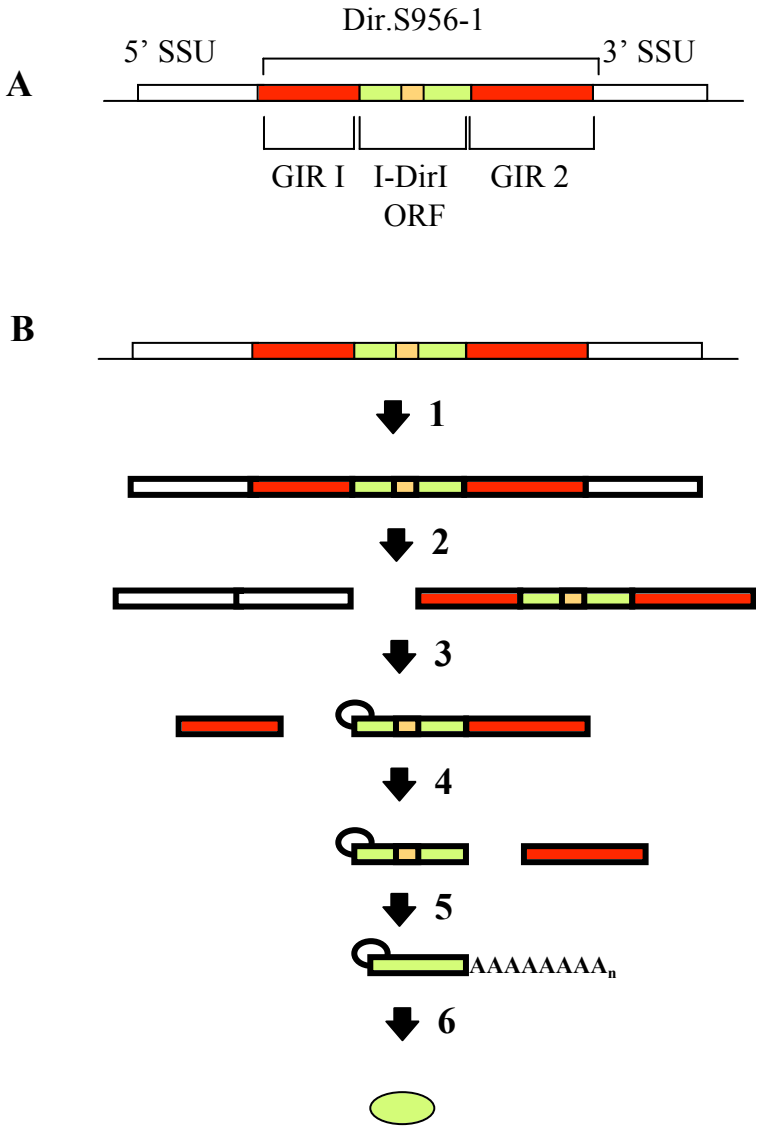


Figure 1.15: Expression of I-DirI from Dir.S956-1. A. A schematic representation of Dir.S956-1. B Expression of I-DirI from Dir.S956-1. 1. The DNA(thin lines) is transcribed by RNA Polymerase I into RNA (thick lines). 2. The GIR2 ribozyme splices the intron RNA out of the exon RNA. 3. The GIR1 ribozyme catalyzes the formation of a 2'-5' lariat at the 5' end of the I-DirI ORF, which separates GIR1 from the rest of the intron. 4-5. The GIR2 ribozyme region is removed, and then most likely the cellular poly-adenylation machinery adds a poly-A tail to the 3' end of the I-DirI ORF. A 51-nt spliceosomal intron (yellow) is removed by the spliceosomal complex. 6. The processed I-DirI RNA is translated into the I-DirI protein.

The homing endonuclease ORF, I-*Dir*I, contains a 51 nucleotide spliceosomal intron (I51). I-DirI appears to be translated from a RNA Pol I transcript which is then processed to mimic a RNA Pol II transcript. After Dir.S956-1 is spliced from the ribosomal small subunit pre-RNA, it undergoes processing to form the 2'-5' lariat, which caps the 5' end and removes the GIR1 sequence, and one cleavage at the 3' end, which removes the majority of the GIR2 sequence from the I-Dir I ORF. The 3' processing cleavage exposes a polyadenylation signal on the 3' end of the I-Dir ORF, which results in the addition of a poly-A tail to the processed ORF. The spliceosomal complex removes the I51 intron, and the processed I-Dir I ORF is exported from the nucleus, where I-*Dir*I is translated (Vader, Nielsen et al. 1999).

Work with the closely related twin-ribozyme group I intron, Nja.S516, demonstrates the importance of the GIR1 cleavage in translation from these introns. After several unsuccessful attempts to trans-integrate the intron into the *S. cerevisiae* genome the intron was transcribed from a plasmid. Nja.S516 is most likely unable to trans-integrate due to either slow splicing of the ribozyme or cleavage of the I-*Nja*I endonuclease at degenerate sites. Translated from a plasmid in yeast, Nja.S516 splices itself from the exon RNA and produces I-*Nja*I protein. The spliced intron produces processing products that are similar in size to those produced by Dir.S956-1. It is likely that the Nja.S516 GIR1 catalyzes a 2'-5' lariat similar to that of Dir.S956-1 but this has not been demonstrated. Mutation of the G binding site in GIR1 prevents processing at IPS1 and IPS2, illustrating that GIR1 is responsible for these cleavages. Additionally, mutant GIR1 introns do not produce any detectable I-*Nja* I activity (Decatur, Johansen et al. 2000).

The intron Dir.S956-2 is found in *Didymium iridis* strain CR8-1, inserted at the same site as Dir.S956-1 (Figure 1.16). Dir.S956-2 encodes a homing endonuclease,

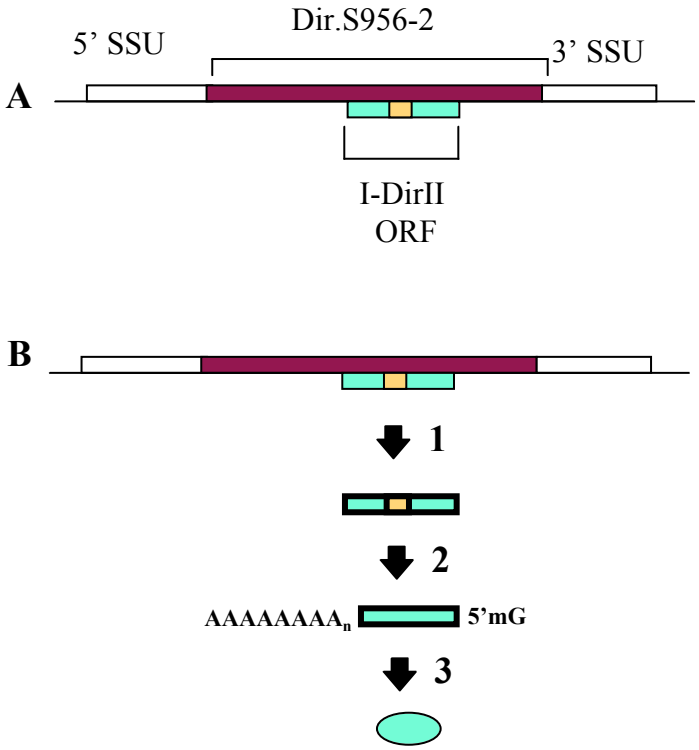


Figure 1.16: Expression of I-DirII from Dir.S956-2. A. A schematic representation of Dir.S956-2. Note that the I-DirII ORF (light blue) is oriented in the antisense direction in relation to Dir.S956-2. B. Expression of I-DirII from Dir.S956-2. 1. The I-DirII gene DNA (thin lines) is transcribed by RNA Polymerase II into RNA (thick lines), independently of the transcription of the SSU and Dir.S956-2. 2. A 50-nt intron is removed by the spliceosomal complex, and a 5' cap and poly-A tail are added to the I-DirII RNA. 3. The processed I-Dir-II RNA is translated into the I-DirII protein.

I-*Dir*II, in its P8 loop. The open reading frame for the I-*Dir*II endonuclease is oriented in the anti-sense direction with respect to Dir.S956-2 and the rest of the small ribosomal subunit gene. The I-*Dir* II message is transcribed by a Pol II like promoter and is polyadenylated at the 3' end. The I-*Dir* II ORF is also interrupted by a 50 nucleotide spliceosomal intron, which was found to be spliced out of the predominance of the polyadenylated ORF RNA (Johansen, Vader et al. 2006).

In summary, very divergent mechanisms for translation of ORFs from nuclear group I introns inserted in rDNA. Although some of these mechanisms have been characterized, many of them are poorly understood.

The role of intron circles

Full-length spliced intron circles have been found in several of the above nuclear group I introns. The roles these structures play, if any, is currently unknown. It is easy to imagine that circular RNA would be more stable than linear RNA, and that the added stability plays a role in the horizontal transfer of the intron (Decatur, Johansen et al. 2000). They could also play a role in translation of intron-encoded proteins, by stabilizing intron RNA or perhaps by mimicking the circles formed by eIFs and poly-A binding protein in mRNA translation initiation (Haugen, De Jonckheere et al. 2002).

Noncanonical translation

There are several examples of translation from non-canonical mRNA. The simplest of these are attempts to express proteins from genes under the control of Pol I and Pol III promoters. These experiments generally have found that proteins can be produced from such transcripts, but at much lower levels than from a Pol II transcript (Fleischer and Grummt 1983; Grummt and Skinner 1985; Smale and Tjian 1985; Lopata, Cleveland et al. 1986; Gunnery and Mathews 1995).

Experiments in *S. cerevisiae* have shown that protein can be translated from an RNA Pol I transcript. The *HIS4* gene was placed under the control of a Pol I promoter and replaced the normal *HIS4* gene on chromosome III. This gene produced RNA that was not capped, but was poly-adenylated, implying that poly-adenylation is not tightly linked to Pol II transcription in yeast. The Pol I derived RNA was transcribed at higher levels than the Pol II transcript, but was also less stable, most likely due to its cap-less state, resulting in the accumulation of similar amounts of the RNA. The Pol I derived RNA was translated inefficiently, at roughly 3% of the level the Pol II derived RNA (Lo, Huang et al. 1998).

Internal ribosome entry sites (IRES) are RNA elements that allow cap-independent translation initiation to occur. These elements allow ribosomes to enter an mRNA internally and allow translation initiation of ORFs downstream of a 5' ORF. The primary assay for these elements is to place a sequence between two reporter genes on a bi-cistronic mRNA. Activity of the second reporter gene indicate IRES activity in the putative IRES sequence. While viral IRES elements have not been shown to function in *S. cerevisiae* without disrupting other players in translation (Thompson, Gulyas et al. 2001), there are currently two characterized examples of naturally occurring IRES elements in the yeast genome, the invasive growth and the URE2 IRES elements. IRES activity has also been demonstrated for yeast genes *HAP4*, *YAP1*, *TAF145*, and *TIF4631 in vivo*, but neither the IRES elements, nor the proteins involved in the internal initiation, have been defined for these genes (Zhou, Edelman et al. 2001; Seino, Yanagida et al. 2005; Raychaudhuri, Fontanes et al. 2006).

As its name implies, the invasive growth IRES is involved in the conversion from non-invasive growth to invasive growth in yeast. This conversion is triggered through glucose starvation (Cullen and Sprague 2000). During the conversion,

translational initiation is inhibited for nearly all genes (Ashe, De Long et al. 2000) except those genes required for invasive growth. All of the invasive growth genes share a short poly-A stretch in their 5' leader sequence. These sequences appear to be responsible for the continued translation of the invasive growth genes. They can confer IRES activity on reporter RNAs electroporated into yeast cells, and deletion of the leader sequences from invasive growth genes prevents their translation during glucose starvation. Chemical probing of the minimal IRES element RNA indicates that the region is unstructured. Translation from this IRES is inhibited by the addition of competing poly-A RNA in yeast translation extracts, and is rescued by the addition of extra Pab1 protein. Thus, it is likely that this IRES functions through interactions with the yeast poly-A binding protein, Pab1, which then allows the recruitment of other translation machinery (Gilbert, Zhou et al. 2007).

There are two forms of the Ure2p naturally found in yeast, both of which are translated from the same mRNA (Komar, Lesnik et al. 2003). The larger of the two proteins is translated from the first AUG on the *URE2* mRNA, while the shorter form of the protein is the result of an initiation at an internal AUG. Analysis of *URE2* truncations placed between a stable stem loop, which inhibits cap-dependent initiation, and a lacZ reporter gene localized the IRES activity to nucleotides 205 to 309 of the *URE2* gene. Structural probing of the region found that the region forms a 25 nucleotide long stem with a 24 nucleotide loop. Interestingly, the first two nucleotides of the IRES-derived Ure2p start codon, nucleotides 279-281, are base paired in this structure. Because of this, the structure must unwind during translation initiation. How this is accomplished, and what factors are involved in initiation at this IRES, are currently unclear (Reineke, Komar et al. 2008).

It is unclear if any of these noncanonical translation pathways is involved in the translation of Ppo.l1925-encoded ORFs in yeast cells. Proteins are translated from

Ppo.L1925 and Ppo-α at levels similar to that of the RNA Pol I-transcribed HIS4 (Lo, Huang et al. 1998; Lin and Vogt 2000). It could be that the intron RNA is translated in a similar fashion. However, this observation does not explain the increases in expression seen in some of the Ppo-α ran1 mutants. The increased level of expression is not caused by an increase in the production or stability of the intron RNA, as the levels of full-length or IPS1- and IPS2-processed intron RNAs are not significantly different between the wild type and high expressing (HEUTR) mutant introns. These mutations could promote interactions between other noncanonical translation pathways, such as the invasive growth IRES pathway, which would increase expression of the intron-encoded ORF. Alternatively, these mutations could promote the formation of a relatively rare intron RNA species, such as a circular intron RNA, which acts as the mRNA for the intron-encoded ORF. This species could have been hidden from detection due to the large amount of full-length and IPS-processed intron RNAs.

Thesis outline

The close relationship of group I introns Ppo.L1925 and Tth.L1925 provides a unique opportunity to study I-PpoI's novel expression from a RNA Pol I transcript. The ribozyme regions of these two introns are very closely related, even though Tth.L1925 lacks a sequence corresponding to the I-PpoI ORF and 3'UTR regions of the Ppo.L1925. Since it does not contain an ORF to be expressed, Tth.L1925 presents itself as a test-bed for understanding how Ppo.L1925 expresses I-PpoI and how Ppo.L1925 sequences affect that expression.

Chapter II: Materials and Methods

The second chapter of this thesis describes the materials and methods used in chapters III-V.

Chapter III: Chimeric ORF-Containing *Tetrahymena* Introns can Trans-Integrate and Produce Protein in *S. cerevisiae.*

It is likely that group I introns have gained homing endonuclease ORFs over the history of their evolution. Since Tth.L1925 and Ppo.L1925 are so similar, can Tth.L1925 introns with ORFs inserted at a site homologous to that of I-PpoI in Ppo.L1925 be trans-integrated into the *S. cerevisiae* genome? Previous work found that larger ORFs, such as Neo and *HIS3*, disrupted the splicing of chimeric Ppo.L1925 introns. Since Tth.L1925 splices more rapidly than Ppo.L1925, will chimeric Tth.L1925 introns with large ORFs, which disrupted the splicing of Ppo.L1925, be able to integrate? If the chimeric Tth.L1925 introns are able to integrate, will their ORFs be expressed and, if they are expressed, at what levels?

In chapter III, I find that chimeric Tth.L1925 introns containing the β-galactosidase α fragment, Neo, or *HIS3* ORFs are able to trans-integrate into the *S. cerevisiae* genome. The proteins encoded by all three of the ORFs are expressed from the chimeric introns. The chimeric Tth.L1925 introns expressed at levels similar to that of the chimeric Ppo.L1925 introns in yeast. Surprisingly, a chimeric Tth.L1925 containing the I-PpoI ORF were unable to integrate into *S. cerevisiae*, possibly due to the lack of the Ppo.L1925 3'UTR sequence.

Chapter IV: The Role of Intron Sequence Elements on expression from Chimeric *Tetrahymena* Introns.

Even though there is over 70% identity in the ribozyme regions of Ppo.L1925 and Tth.L1925, there are regions of the introns that differ. The Ppo.L1925 3' UTR sequence has been shown to have a large effect on both intron splicing and expression levels of intron encoded ORFs. Deletion of the 3' UTR prevents the Ppo.L1925 intron for integrating into *S. cerevisiae* and changing the sequence of the last seven nucleotides can decrease or vastly increase protein expression from the intron. How

would adding the 3'UTR to chimeric Tth.L1925 affect expression? Would the HE 3' URT sequences that increase expression in Ppo.L1925 have the same effect on expression from Tth.L1925? The 5' ends of the two introns are also very different. The 5' end of the intron RNA could play a role in translation initiation. How would the replacing the 5' end of Tth.L1925 with that of Ppo.L1925 affect expression?

In chapter IV, I investigate how the Ppo.L1925 3' UTR and 5' end affect translation from chimeric Tth.L1925 introns. I used the chimeric Tth.L1925 intron containing the α fragment of β-galactosidase ORF, Tth-α, for these experiments, since expression levels are easily measured through β-galactosidase activity. Addition of the Ppo.L1925 3' UTR to Tth-α slightly decreased β-galactosidase activity from the intron. Mutation of the last seven nucleotides of the 3' UTR to the Ppo.L1925 HE sequences did not increase expression from Tth-α. Mutating the 5' end of Tth-α to that of Ppo.L1925, to create the intron Tth-α-5'P, increased β-galactosidase activity ten fold. Adding the Ppo.L1925 3'UTR to Tth-α-5'P affected expression from the intron in a similar fashion to the how they affected expression from Tth-α. Both the wild type and HE 3'UTRs reduced expression when added to Tth-α-5'P suggesting that the HE 3'UTRs are not interacting with the 5' end of Ppo.L1925.

In order to gain a better understanding of the role of the intron's 5' end, I created two pools of introns with randomized sequence at the 5' end of the intron. The first pool randomized four nucleotides in the 5' end of Tth-α. Many of these introns were able to integrate into the yeast genome, but none of them exhibited a significant increase in β-galactosidase over Tth-α. The second pool contained randomized nucleotides at eight positions in the longer 5' end of Tth-α-5'P. As with the previous pool, a large percentage of the introns from this pool were able to integrate. Roughly half of the introns tested have similar activity to Tth-α-5'P, one quarter had activity similar to Tth-α, and one quarter had little to no activity. There does not seem to be a

consensus sequence that increased expression in the intron. These results indicate that, while changes in the sequence identity of the intron 5' end has little to no effect on intron splicing, it does have a large effect on expression of the intron encoded ORF.

Chapter V: Investigation of the Mechanism of Translation of Chimeric *Tetrahymena* Intron-Encoded ORFs

There are currently two examples of cap-independent translation in wild type *S. cerevisiae*, the glucose-starvation IRES sequences and the URE2 IRES. The URE2 IRES is unlikely to be involved in expression from Ppo.L1925 and the chimeric Tth.L1925 introns, since it is so large. The glucose-starvation IRES, on the other hand, is a small sequence and it seemed possible that the chimeric-intron encoded proteins are expressed through this pathway. Alternatively, the intron RNA circles could be acting as the message for translation of the intron-encoded ORFs. Could either of these be factors in intron encoded ORF translation?

In chapter V, I investigate the affect of the glucose starvation IRES sequence on expression from chimeric Tth.L1925. The adition of a glucose starvation IRES sequence to a chimeric intron did not signifigantly increase the amount of expression from the intron. This leads to the conclusion that the Tth.L1925 encoded ORFs are not expressed through the glucose starvation ORF pathway. I also examined the role of intron RNA circles in expression. I examined chimeric Tth.L1925 and Ppo.L1925 introns for the formation of intron RNA circles. All of the introns tested formed either full-length or smaller circles. However, there is no apparent correlation between the junction of the circular RNAs and the expression level of the intron encoded ORF, which suggests that the circular introns are not acting as the message.

Chapter VI: Perspectives on Future Research.

In chapter VI, I discuss other models for the expression of ORFs encoded by chimeric group I introns and how understanding how these ORFs are expressed could provide us with a novel tool to produce large amounts of protein *in vivo*.

Chapter II

Materials and Methods

Plasmid Construction

Plasmids used are listed in Table 2.1. Plasmids were constructed by standard cloning methods (Sambrook, Fritsch et al. 1989). Mutations were introduced by multiple-step PCR with mutations included in the primers (see Table 2.2 for primer sequences) as described (Vallett, Merge et al. 1988). Two external primers with the appropriate restriction enzyme sites and complementary internal primers containing the desired mutations were used in this process. Two parallel PCR reactions, each using one external primer and one mutagenic primer, were performed. The products from the first step were gel-purified with the QIAEXII kit (Qiagen, Catalog number 20051). The purified PCR products were then combined and used as templates with the two external primers to create the final PCR product. The PCR product was purified using the High Pure PCR Purification Kit (Roche, Catalog number 11732668001). It was then cloned into the desired vector using standard methods (Sambrook, Fritsch et al. 1989).

pJLTth-ClaI was made by amplifying the 5' end of the Tth.L1925 intron and upstream yeast rDNA with JL9 and JL165 and the 3' half of the Tth.L1925 with JL164 and JL85. The resulting two products were combined and amplified using JL9 and JL85 to make the full-length Tth-ClaI intron flanked by rDNA sequence. This product was cloned into EcoRI-SalI digested pRS423 (Christianson, Sikorski et al. 1992).

Table 2.1: Plasmids used.

Plasmid Name	Backbone	Additional sequence	Other Names	References
pRS423	n.a.			Christianson, Sikorski et al., 1992
pCpIPpo	pYES	Gal 1,10 driven I-PpoI ORF		Lin and Vogt, 1998
pRSTth-ClaI	pRS423	Intron Tth-ClaI		This work
pRSTth-α	pRS423	Intron Tth-α	pRSTtLSU1-α	This work
pRSTth-Neo	pRS423	Intron Tth-Neo	pRSTtLSU1-Neo	This work
pRSTth-His3HA	pRS423	Intron Tth-His3HA		This work
pRSTth-IPpoHA	pRS423	Intron Tth-IPpoHA		This work
pRSPpo-αHEUTR A	pRS423	Intron Ppo-αHEUTR A	pRSI3αran1 9-7	This work
pRSPpo-αHEUTR B	pRS423	Intron Ppo-αHEUTR B	pRSI3αran1 13-47	This work
pRSPpo-αHEUTR C	pRS423	Intron Ppo-αHEUTR C	pRSI3αran1 10-2	This work
pRSPpo-αHEUTR D	pRS423	Intron Ppo-αHEUTR D	pRSI3αran1 2-26	This work
pRSPpo-αHEUTR E	pRS423	Intron Ppo-αHEUTR E	pRSI3αran1 10-11	This work
pRSPpo-αHEUTR F	pRS423	Intron Ppo-αHEUTR F	pRSI3αran1 12-9	This work
pRSTth-αUTR	pRS423	Intron Tth-αUTR	pRSTtLSU1-αUTR	This work
pRSTth-αHEUTR A	pRS423	Intron Tth-αHEUTR A	pRSTtLSU-αHEUTR A	This work
pRSTth-αHEUTR B	pRS423	Intron Tth-αHEUTR B	pRSTtLSU1-αHEUTR B	This work
pRSTth-αHEUTR C	pRS423	Intron Tth-αHEUTR C	pRSTtLSU1-αHEUTR C	This work
pRSTth-5’P-α	pRS423	Intron Tth-5’P-α	pRSTtLSU1P-α 5'P	This work
pRSTth-5’P-αUTR	pRS423	Intron Tth-5’P-αUTR	pRSTtLSU1-αUTR 5'P	This work
pRSTth-5’P-αHEUTR A	pRS423	Intron Tth-5’P-αHEUTR A	pRSTtLSU1-αHEUTR A 5’P	This work

Table 2.1 (Continued)

Plasmid Name	Backbone	Additional sequence	Other Names	References
pRSTth-5'P-αHEUTR B	pRS423	Intron Tth-5'P-αHEUTR B	pRSTtLSU1-αHEUTR B 5'P	This work
pRSTth-5'P-αHEUTR C	pRS423	Intron Tth-5'P-αHEUTR C	pRSTth-αHEUTR C 5'P	This work
pRSTth-5'P-His3HA	pRS423	Intron Tth-5'P-His3HA		This work
pRSTth-αHA	pRS423	Intron Tth-αHA		This work
pRSTth-IRESαHA	pRS423	Intron Tth-IRESαHA		This work
pRSTth-RCIRESαHA	pRS423	Intron Tth-RCIRESαHA		This work

Table 2.2: Primer sequences.

Name	Primer Sequence (5' to 3')
JL8	TATATCGATTCTGCCAAGCCCGT
JL9	CGTGAATTCAACTTAGAACTGGTAC
JL83	TCACCCCGGAATTGGTTTATCC
JL84	CGAATGGGACCTTGAATGC
JL85	ACAGGTCGACAAGGTAGTGGTATTTCACTGG
JL164	CTAAATAGCATCGATTACCTTTGGAGGGAAAAG
JL165	CCAAAGGTAATCGATGCTATTTAGAGAGTC
PrRS31	ATCATCGATATGATTGAACAAGATGGATTGC
PrRS32	ATCATCGATTCAGAAGAACTCGTCAAGAAGG
PrRS35	ATCATCGATCGGTCAGCGTTTTGCCCGTGGCGCTGACCCACCCGATCTAGTAACCTC
PrRS36	ATCATCGATCCTATTGCGTTTTGCCCGTGGCGCTGACCCACCCGATCTAGTAACCTC
PrRS39	ATCATCGATATGGCGCTCACCAATCGCTTAATTACGG
PrRS40	ATCATCGATTTATACCACAAAGTGACTAATTGATAAGACG
PrRS41	CCCGATCTAGTAACCTCAGACCTTATACCACAAAGTGACTAA
PrRS42	GGATCCGATATCCCCGGGCTGCAGGAATTCAAC
PrRS43	TGGGATCCGGGCCCCCCCTCGAGGTCG
PrRS44	GTAATCGATGCTATTTAGAGAGTCATAGTTACTCCCGCCGTTTACTCGCGCTTG
PrRS49	TAATACGACTCACTATAGGGCGAGTACTCCAAAAC
PrRS53	ATCGATTACCGTAGGAGAGAAAAG
PrRS54	CTTTTCTCTCCTACGGTAATCG
PrRS60	GACGGTCTTGCCTTTTAAACCG
PrRS61	ATCTCTAGAATGGCGCTCACCAATCGCTTAATTACGG
PrRS62	ATCGTCGACTTATACCACAAAGTGACTAATTG
PrRS65	CCCTCACTAAAGGGAACAAAAGCTGG
PrRS66	CGACTCACTATAGGGCGAATTGGG
PrRS71	GCTGCAAGGCGATTAAGTTGGG
PrRS72	CCCTTAATGGGGACCTGGAGAAG
PrRS73	GCGCGAGTAAACGGCGGGAGTAACTATGACTCTCTCACCCCCTTAAATCGATATGGCGCTCACCAATCGC
PrRS74	AGAGAGTCATAGTTACTCCCGCCGTTTACTC
PrRS79	TAATACGACTCACTATAGGGCGCTGATTCCAAACTCGGGTGC
PrRS80	ACCAAGTAGGAGAGAAAAGTCAC
PrRS81	TAATACGACTCACTATAGGGTTATACCACAAAGTGACTAATCG
PrRS82	CACCTGGTAGCTAGTCTTTAAACC
PrRS83	TAATACGACTCACTATAGGGATGGCGCTCACCAATCG

Table 2.2 (Continued)

Name	Primer Sequence (5' to 3')
PrRS85	CGAGTACTCCAAAAC
PrRS86	CGACGGCCAGTGAATCCGTAATTAAGCGATTGGTGAGCGCCAT
PrRS87	CGAGTACTCCAAAACTAATCAATATACTTTCGCATACAAATTAG
PrRS88	ATCTTCGAATTATACCACAAAGTGACTAATTGATAAGACG
PrRS89	GATATCATCGATNNNNNNNNGTGAGAGAGTCATAGTTACTCCCGC
PrRS90	ATCGAATTCTTCCATAGGCTTCCGCCCCCCTGACGAGCATC
PrRS91	CGGGCTTCCCATACAATCTGTAGATTGTCGCACCTGATTGCCC
PrRS92	GGGCAATCAGGTGCGACAATCTACAGATTGTATGGGAAGCCCG
PrRS93	GATATCATCGATTTAGAAAAACTCATCGAGCATCAAATGAAAC
PrRS97	GGGTCAGCGCCACGGGCAAAACGCGTTTAAATCGATTACC
PrRS98	GGGTCAGCGCCACGGGCAAAACGCCCTCGAATCGATTACC
PrRS101	GATATCATCGATNNNNTTTAGAGAGTCATAGTTACTCCCGC
PrRS103	AGTGATGCAACACTGGAGCCGCTGGG
PrRS104	TCACCCCGGAATTGGTTTATCCGGAG
PrRS105	CGAATGGGACCTTGAATGCTAGAACG
PrRS109	TAATACGACTCACTATAGGGCGAAAAAATCGGACCGGCCAACC
PrRS110	CGAAAAAATCGGACCGGCCAACC
PrRS111	TCGATTCCGGAGAGGGAGC
PrRS112	ATCATCGATATGACAGAGCAGAAAGCCC
PrRS113	ATCATCGATCTACATAAGAACACCTTTGGTGG
PrRS114	ATCATCGATATGGCGCTCACCAATGCTC
PrRS115	ATCATCGATTTATACCACAAAGTGACTGCCC
PrRS119	ATCATCGATTATGCATAGTCCGGGACGTCATAGGGATAGCCCGCA TAGTCAGGAACATCGTATGGGTA
PrRS120	GGAACATCGTATGGGTATACCACAAAGTGACTGCCC
PrRS123	GGAACATCGTATGGGTATACCACAAAGTGACTAATTGATAAGACG
PrRS127	ATCATCGATAACAAAAAAAAAATGGCGCTCACCAATCGCTTAATT ACGG
PrRS128	ATCATCGATTTTTTTTTTGTTATGGCGCTCACCAATCGCTTAATTAC GG
PrRS157	CCCTTTCCCGCAATTTGACGGTC
PrRS158	CCAGGTGCCTACCTGTGAC

pRSTth-ClaI was made to correct a C to T mutation in at the -2 position in pJLTth-ClaI. A fragment was amplified using PrRS42 and PrRS44. This fragment was then cloned into the BamHI and ClaI sites in pJLTth-ClaI. PrRS42 adds an EcoRV site to the 5' MCS that was not in pJLTth-ClaI.

pRSTth-α was made by amplifying the α fragment from $pJLI_3\alpha$ with prRS39 and prRS40 which was then cloned into the ClaI site in pRSTth-ClaI.

pRSTth-Neo was made by amplifying the Neo fragment from pET28-b+ with PrRS31 and prRS32 which was then cloned into the ClaI site in pRSTth-ClaI.

pRSTth-His3HA was made by amplifying the *HIS3* gene from plasmid pRS423 with primers PrRS112 and PrRS113. The resulting PCR product was amplified using PrRS112 and PrRS119 to added the HA tag to the 3' end. The product of the second amplification was cloned into Tth-ClaI

pRSTth-IPpoHA was made by amplifying the I-PpoI gene from plasmid pJLI3 with primers PrRS114 and PrRS115. The resulting PCR product was amplified using PrRS114 and PrRS119 to added the HA tag to the 3' end. The product of the second amplification was cloned into Tth-ClaI

pRSPpo-αHETUR A through F were made by amplifying the appropriate intron from yeast total genomic DNA using primers JL9 and JL85 and cloning the amplified product into EcoRI-SalI cut pRS423.

pRSTth-αUTR was made by amplifying the α fragment from pRSTth-α with PrRS39 and prRS41 and then re-amplifying with PrRS39 and prRS35 to add the 3' end of the UTR and the ClaI site. The second round PCR product was then cloned into the ClaI site in pRSTth-ClaI.

pRSTth-αHEUTR A was made by amplifying the α fragment from pRSTth-α with PrRS39 and PrRS41 and then re-amplifying with PrRS39 and PrRS36 to add the

3’ end of the UTR and the ClaI site. The second round PCR product was then cloned into the ClaI site in pRSTth-ClaI.

pRSTth-αHEUTR B was made by amplifying the α fragment from pRSTth-α with PrRS39 and PrRS41 and then re-amplifying with PrRS39 and PrRS97 to add the 3’ end of the UTR and the ClaI site. The second round PCR product was then cloned into the ClaI site in pRSTth-ClaI.

pRSTth-αHEUTRC was made by amplifying the α fragment from pRSTth-α with PrRS39 and PrRS41 and then re-amplifying with PrRS39 and PrRS98 to add the 3’ end of the UTR and the ClaI site. The second round PCR product was then cloned into the ClaI site in pRSTth-ClaI.

pRSTth-5’P-ClaI was made through a three step process. The 5’ insert piece was made by amplifying the region between the EcoRI site and ClaI site of pRSTth-ClaI with PrRS65 and PrRS74. The middle insert piece was made by amplifying the region between the ClaI site and the pRSTth-ClaI P1’ region with primers PrRS73 and PrRS54. The 3’ insert piece was made by amplifying the region between the P1’ region and the XhoI site of pRSTth-ClaI with PrRS53 and PrRS66. A fragment fusing the 5’ piece and middle piece was made by combining the 5’ and middle insert pieces and amplifying using PrRS65 and PrRS54. The 5’-middle piece was then fused to the 3’ piece by combining them and amplifying with PrRS65 and PrRS66. This 5’-middle-3’ fusion piece was then cloned between the EcoRI and SphI sites in pRSTth-ClaI to make pRSTth-5’P-ClaI.

pRSTth-5’P-α was made by inserting the α fragment from ClaI-digested pRSTth-α into the ClaI site of pRSTth-5'P-ClaI.

pRSTth-5’P-αUTR was made by amplifying the α-UTR fragment from pRSTth-αUTR with primers PrRS39 and PrRS35 and cloning it into the ClaI site of

pRSTth-ClaI. This amplification was necessary since the ClaI site 3' of the αUTR fragment of pRSTth-αUTR is methylated by Dam methylase.

pRSTth-5'P-αHEUTR A was made by amplifying the α-UTR fragment from pRSTth-αHEUTR A with primers PrRS39 and PrRS36 and cloning it into the ClaI site of pRSTth-ClaI. (This amplification was necessary since the ClaI site 3' of the αUTR fragment of pRSTth-αUTR is methylated by Dam methylase.)

pRSTth-5'P-αHEUTR was made by amplifying the α-UTR fragment from pRSTth-αHEUTR B with primers PrRS39 and PrRS97 and cloning it into the ClaI site of pRSTth-ClaI.

pRSTth-5'P-αHEUTR C was made by amplifying the α-UTR fragment from pRSTth-αHEUTR B with primers PrRS39 and PrRS98 and cloning it into the ClaI site of pRSTth-ClaI.

pRSTth-5'P-His3HA was made by cloning the *HIS3*-HA fragment from pRSTth-His3HA into the ClaI site of pRSTth-5'P-ClaI.

pRSTth-αHA was made by amplifying the α fragment from pRSTth-α with primers PrRS39 and PrRS123. The product of this amplification was re-amplified using PrRS39 and PrRS119 to add the duel HA tags to the sequence. The αHA fragment was then cloned into the ClaI site of pRSTth-ClaI.

pRSTth-IRES-αHA was made by amplifying the αfragment from pRSTth-α with primers PrRS127 and PrRS123. The product of this amplification was re-amplified using PrRS127 and PrRS119 to add the duel HA tags to the sequence. The αHA fragment was then cloned into the ClaI site of pRSTth-ClaI.

pRSTth-RCIRES-αHA was made by amplifying the α fragment from pRSTth-α with primers PrRS128 and PrRS123. The product of this amplification was re-amplified using PrRS127 and PrRS119 to add the duel HA tags to the sequence. The αHA fragment was then cloned into the ClaI site of pRSTth-ClaI.

Randomization

To make the vector for the 5' end randomization in the context of the Tth-5'P-α intron (a.k.a. the Tth-5'P-α-ran pool), pRSTth-5'P-α-BST-KAN (Figure 2.1), the α fragment was amplified from pRSTth-α using primers PrRS39 and PrRS88. This replaced the 3' ClaI site with a BstBI site. This amplified fragment was cloned into the ClaI site of Tth-5'P-ClaI. The resulting plasmid, pRSTth-5'P-α-BST, has a ClaI site 5' of the α gene and a hybrid ClaI/BstBI site 3' of the α gene. To remove the internal ClaI site from the KAN gene, it was amplified in two stages. The first half of the KAN gene was amplified from plasmid pET28-b+ using primers PrRS90 and PrRS91. The second half of the KAN gene was amplified using and PrRS93. The two KAN halves were then combined and amplified using PrRS90 and PrRS93 to make a full-length KAN fragment. The resulting KAN fragment was cloned between the EcoRI and ClaI sites of pRSTth-5'P-α-BST to make pRSTth-5'P-α-BST-KAN. The gel-purified EcoRI-ClaI-cut pRSTth-5'P-α-BST-KAN backbone was then gel purified

To make the randomized pool, the region between the EcoRI site and ClaI site of pRSTth-5'P-α was amplified using primers PrRS65 and PrRS89. PrRS89 contains equal concentrations of A, C, G, and T at the N positions. This pool was digested with EcoRI and ClaI and gel purified. To make the plasmid pool, 300 ng of digested vector was combined with 250 ng of digested insert and ligated at 16˚C overnight. The ligation mixture was ethanol precipitated, washed twice with 1 ml of cold 70% ethanol, and transformed by electroporaton (BioRad GenePulser II, 2.0 kV, 200 W, 25 μF, 0.1 cm electroporation cuvette) into ElectroMAX DH5a-E cells (Invitrogen Catalog number 11319019). To estimate the number of transformants, 1%, 0.1%, and 0.01% of the transformed *E. coli* were plated on 2YT, 2YT + Amp (to estimate pool size) and 2YT + Amp +Kan plates (to estimate the amount of uncut plasmid). The remainder of the transformation was added to 200 ml of

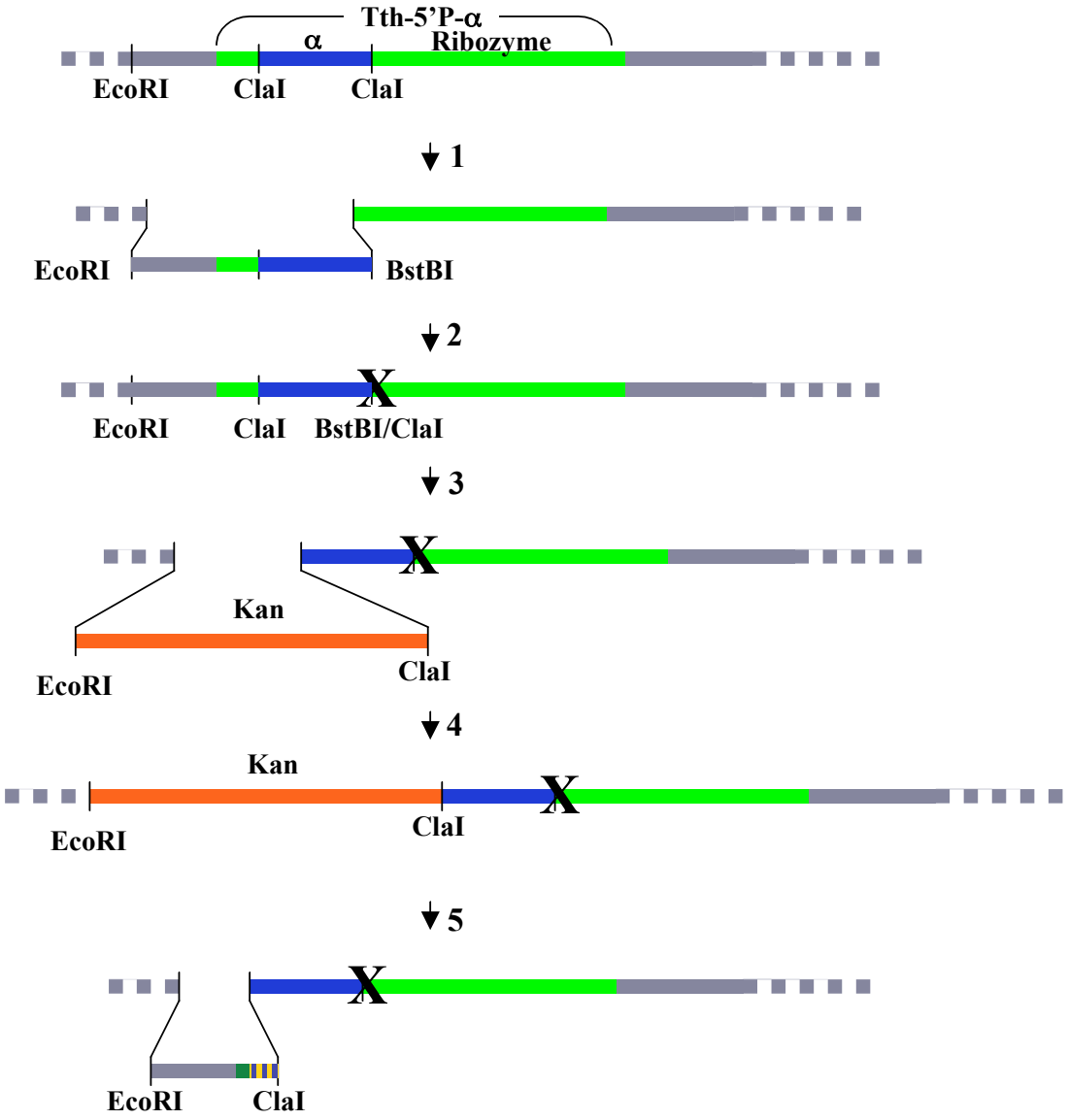


Figure 2.1: Construction of pRSTth-5'P-α-BST-KAN. 1. A PCR product consisting of the 5' LSU sequence, 5' intron sequence, and alpha ORF with the 3' ClaI site replaced with a BstBI site is cloned into EcoRI and ClaI digested pRSTth-5'P-α. 2. The ClaI/BstBI hybrid site disrupts the 3' ClaI site. 3-4. The resulting plasmid is again digested with EcoRI and ClaI, and a Kan stuffer fragment is cloned into it. 5. The Kan stuffer fragment is removed by digesting with EcoRI and ClaI and a replaced with a fragment from a pool with randomized nucleotides at the intron 5' end.

liquid 2YT + Amp and incubated at 37˚C for 16 hours to increase the amount of DNA in the pool. The plasmid DNA was harvested from the amplified pool using a Qiagen Midi Kit (Catalog number 12143) and co-transformed into yeast strain NOY505GalΩT with plasmid pCPIPpo. To estimate transformation efficiency, 1% of the transformed yeast was plated on SD plates. The remainder of the transformation mixture was plated on SD–Ura-His plates. Yeast transformants were washed off of the plates and combined. A fraction of the pool (5%) was grown in 100 ml of SD-Ura-His over night. Cells were collected, washed twice with SG-Ura-His media, and then plated on SG-Ura-His plates. Individual colonies were selected from the SG-Ura-His plates and patched on to SG+XGal plates to screen for β-galactosidase activity.

The vector for the Tth-α-context 5' end randomization (the Tth-α-ran pool), pRSTth-α-BST-KAN, was made as above, except that pRSTth-α was used as a starting point instead of pRSTth-5'P-α. Primers PrRS65 and PrRS101 were used to make the randomized pool. PrRS101 contains equal concentrations of A, C, G, and T at the N positions.

Yeast strains and media

Yeast strains used are listed in Table 2.3. Yeast were grown according to procedures in (Sherman, Fink et al. 1986). Yeast cultures were grown in YEPD (1% yeast extract, 2% Bacto-peptone, 2% dextrose) or, when induction of Gal-promoter genes was necessary, YEPG (1% yeast extract, 2% Bacto-peptone, 2% galactose). When selection was needed, yeast cells were grown in synthetic minimal medium (S medium) (0.17% yeast nitrogen base (without amino acids or ammonium sulfate), 0.5% ammonium sulfate) supplemented with amino acids as needed (30 mg/ml of L-histidine, 40 mg/ml of L-leucine, 30 mg/ml of L-Lysine, and 30 mg/ml of L-Tryptophan).

Table 2.3: Yeast strains used.

Strain	Genotype	Integrated Intron	Notes	Source
INVSc2	MATA his3-D200 ura3-167	None		Invitrogen, Inc.
INVSc2/Tth-ClaI	similar to INVSc2	Tth-ClaI		This work
INVSc2/Tth-Neo	similar to INVSc2	Tth-Neo		This work
Noy505GalΩT	MATa ade2 ura3 leu2 trp1 his3 can1	None	β-gal Ω under gal 1,10 promoter control in trp1 locus	Lin and Vogt, 2000
Noy505GalΩT/Tth-ClaI	similar to Noy505GalΩT	Tth-ClaI		This work
Noy505GalΩT/Tth-α	similar to Noy505GalΩT	Tth-A		This work
Noy505GalΩT/Tth-His3HA	similar to Noy505GalΩT	Tth-His3HA		This work
Noy505GalΩT/Ppo-αHEUTR A	similar to Noy505GalΩT	Ppo-αHEUTR A		This work
Noy505GalWT/Ppo-αHEUTR B	similar to Noy505GalΩT	Ppo-αHEUTR B		This work
Noy505GalWT/Ppo-αHEUTR C	similar to Noy505GalΩT	Ppo-αHEUTR C		This work
Noy505GalWT/Ppo-αHEUTR D	similar to Noy505GalΩT	Ppo-αHEUTR D		This work
Noy505GalWT/Ppo-αHEUTR E	similar to Noy505GalΩT	Ppo-αHEUTR E		This work
Noy505GalWT/Ppo-αHEUTR F	similar to Noy505GalΩT	Ppo-αHEUTR F		This work
Noy505GalΩT/Tth-αUTR	similar to Noy505GalΩT	Tth-αUTR		This work
Noy505GalΩT/Tth-αHEUTR A	similar to Noy505GalΩT	Tth-αHEUTR A		This work
Noy505GalΩT/Tth-αHEUTR B	similar to Noy505GalΩT	Tth-αHEUTR B		This work

Table 2.3 (Continued)

Strain	Genotype	Integrated Intron	Notes	Source
Noy505GalΩT/Tth-αHEUTR C	similar to Noy505GalΩT	Tth-αHEUTR C		This work
Noy505GalΩT/Tth-5'P-α	similar to Noy505GalΩT	TtLSU1-5'P-α		This work
Noy505GalΩT/Tth-5'P-αUTR	similar to Noy505GalΩT	Tth-5'P-αUTR		This work
Noy505GalΩT/Tth-5'P-αHEUTR A	similar to Noy505GalΩT	Tth-5'P-αHEUTR A		This work
Noy505GalΩT/Tth-5'P-AHEUTR B	similar to Noy505GalΩT	Tth-5'P-αHEUTR B		This work
Noy505GalΩT/Tth-5'P-αHEUTR C	similar to Noy505GalΩT	Tth-5'P-αHEUTR C		This work
Noy505GalΩT/Tth-5'P-α-ran (1-18)	similar to Noy505GalΩT	Tth-5'P-α-ran (1-18)*	18 different strains see Figure 4.15	This work
Noy505GalΩT/Tth-α-ran (1-24)	similar to Noy505GalΩT	Tth-α-ran (1-24)*	24 different strains see Figure 4.18	This work
Noy505GalΩT/Tth-AHA	similar to Noy505GalΩT	Tth-αHA		This work
Noy505GalΩT/Tth-IRESαHA	similar to Noy505GalΩT	Tth-IRESαHA		This work
Noy505GalΩT/Tth-5'RCIRESαHA	similar to Noy505GalΩT	Tth-5'RCIRESαHA		This work

Trans-integration

Two plasmids, pCPIPpo and a plasmid (Table 2.1) encoding the intron of interest (Table 2.4), were co-transformed into the yeast strain INVSc2 (Invitrogen Corporation) or NOY505GalΩT (Lin and Vogt 2000) either by the lithium acetate method (Ausubel, Brent et al. 1990) or using the Frozen-EZ Yeast Transformation II (Zymo Research, Catalog number T2001)**.** To select for yeast that had taken up both plasmids, transformations were plated on SD-Ura-His plates. Transformants were streaked on SGal-Ura-His plates to induce the expression of I-PpoI from pCPIPpo. Single colonies on SGal-Ura-His plates were grown in SGal-Ura-His liquid medium. Genomic DNA was prepared from the above culture and PCR was performed to screen for intron-integrated colonies. Primer pair JL83 and JL84 or PrRS104 and PrRS105, both spanning the Ppo.L1925 and Tth.L1925 insertion site, were used for this PCR analysis. The cells were cured of plasmids by growing them in liquid YEPD for approximately 10 generations and plating on YEPD plates. The loss of the plasmids was checked by replicating colonies from YEPD plates onto SD-Ura and SD-His plates.

Yeast genomic DNA preparation

Yeast cultures were grown to an OD_{600} of 1-3 in YEPD or YEPG media. Cells from 3 ml of culture were pelleted by centrifugation and then resuspended in 200 ml of breaking buffer (2% Triton X-100, 1% SDS, 100mM NaCl, 10mM Tris HCl [pH 8.0], 1mM EDTA [pH 8.0]). Approximately 200 µl of sterile glass beads (425-600 micron, Sigma) and 200 ml phenol/chloroform (equilibrated with TE) was added to the cell suspension. The mixture was then vortexed at highest speed for 3 min to break the yeast cells. Two hundred µl of TE (pH 8.0) was added and the cells were spun at maximum speed in a tabletop centrifuge for 5 min at room temperature. The aqueous layer was transferred to a fresh tube and the DNA was precipitated by adding

Table 2.4: Introns used.

*See figure 4.14 and 4.18 for Tth-5'P-A-ran and Tth-α-ran 5' end sequences.

Intron Name	Ribozyme	ORF	Other Sequence	Other Names
Ppo.L1925	Ppo.L1925	I-PpoI	3'UTR	PpLSU3, I3
Ppo-α	Ppo.L1925	α	3'UTR	JLI3-α
Ppo-αHEUTR A	Ppo.L1925	α	I3αran1 9-7	9-45, 10-12
Ppo-αHEUTR B	Ppo.L1925	α	I3αran1 13-47	
Ppo-αHEUTR C	Ppo.L1925	α	I3αran1 10-2	
Ppo-αHEUTR D	Ppo.L1925	α	I3αran1 2-26	9-18
Ppo-αHEUTR E	Ppo.L1925	α	I3αran1 10-11	
Ppo-αHEUTR F	Ppo.L1925	α	I3αran1 12-9	
Tth.L1925	Tth.L1925	n/a	n/a	TtLSU1
Tth-ClaI	Tth.L1925	n/a	ClaI in P1	TtLSU1-ClaI
Tth-α	Tth.L1925	α	n/a	TtLSU1-α
Tth-Neo	Tth.L1925	Neo	n/a	TtLSU1-Neo
Tth-His3HA	Tth.L1925	His3HA	n/a	n/a
Tth-IPpoHA	Tth.L1925	I-PpoIHA	n/a	n/a
Tth-αUTR	Tth.L1925	α	3'UTR	TtLSU1-αWTUTR
Tth-αHEUTR A	Tth.L1925	α	3'HEUTR 9-7	TtLSU1-αHEUTRA
Tth-αHEUTRB	Tth.L1925	α	3'HEUTR 13-47	TtLSU1-αHEUTRB
Tth-αHEUTRC	Tth.L1925	α	3'HEUTR 10-2	TtLSU1-αHEUTRC
Tth-5'P-α	Tth.L1925	α	Ppo.l1925 5' end	TtLSU1-α5'P
Tth-5'P-His3HA	Tth.L1925	His3HA	Ppo.l1925 5' end	n/a
Tth-5'P-αUTR	Tth.L1925	α	Ppo.l1925 5' end, 3'UTR	TtLSU1-α5'PWTUTR
Tth-5'P-αHEUTR A	Tth.L1925	α	Ppo.l1925 5' end, HEUTR 9-7	TtLSU1-A5'PHEUTRA

Table 2.4 (Continued)

Intron Name	Ribozyme	ORF	Other Sequence	Other Names
Tth-5'P-αHEUTR B	Tth.L1925	α	Ppo.l1925 5' end, HEUTR 13-47	TtLSU1-α5'PHEUTRB
Tth-5'P-αHEUTR C	Tth.L1925	α	Ppo.l1925 5' end, HEUTR 10-2	TtLSU1-α5'PHEUTRC
Tth-αHA	Tth.L1925	αHA	n/a	n/a
Tth-IRESαHA	Tth.L1925	αHA	YMR181c IRES	n/a
Tth-RCIRESαHA	Tth.L1925	αHA	Reverse complement of YMR181c IRES	n/a
Tth-5'α-ran 1	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 D3, D4, D5, D8
Tth-5'α-ran 2	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 D11
Tth-5'α-ran 3	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 D7
Tth-5'α-ran 4	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 D1
Tth-5'α-ran 5	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 L7
Tth-5'α-ran 6	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 L11
Tth-5'α-ran 7	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 D2
Tth-5'α-ran 8	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 D12
Tth-5'α-ran 9	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 D9
Tth-5'α-ran 10	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 D6
Tth-5'α-ran 11	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 L5
Tth-5'α-ran 12	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 L8
Tth-5'α-ran 13	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 L10
Tth-5'α-ran 14	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 L2
Tth-5'α-ran 15	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 L3
Tth-5'α-ran 16	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 L4
Tth-5'α-ran 17	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 L9

Table 2.4 (Continued)

Intron Name	Ribozyme	ORF	Other Sequence	Other Names
Tth-5'α-ran 18	Tth.L1925	α	Ppo.l1925 5' end, 8 nt randomized*	ran89 L6
Tth-α-ran 1	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 D2
Tth-α-ran 2	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 D3
Tth-α-ran 3	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 D5
Tth-α-ran 4	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 D8
Tth-α-ran 5	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 D11
Tth-α-ran 6	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 L1
Tth-α-ran 7	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 L2
Tth-α-ran 8	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 L5
Tth-α-ran 9	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 D1
Tth-α-ran 10	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 D4
Tth-α-ran 11	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 D6
Tth-α-ran 12	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 D7
Tth-α-ran 13	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 D9
Tth-α-ran 14	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 D10
Tth-α-ran 15	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 D12
Tth-α-ran 16	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 L3
Tth-α-ran 17	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 L4
Tth-α-ran 18	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 L6
Tth-α-ran 19	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 L7

Table 2.4 (Continued)

Intron Name	Ribozyme	ORF	Other Sequence	Other Names
Tth-α-ran 20	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 L8
Tth-α-ran 21	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 L9
Tth-α-ran 22	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 L10
Tth-α-ran 23	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 L11
Tth-α-ran 24	Tth.L1925	α	Tth.L1925 5' end, 4 nt randomized*	ran101 L12

1 ml of 100% ethanol. DNA was spun down at room temperature for 3 minutes and resuspended in 400 ml of TE. To remove they yeast RNA, 6 µl of 10 mg/ml RNase A was added and the DNA was incubated at 37°C for 5 minutes. DNA was precipitated by adding 10 ml of 5M NH_4OAC and 1 ml of 100% ethanol and resuspended in 50 ml TE (pH8.0). Two µl of DNA was used for a 50 ml PCR reaction.

RNA preparation

RNA was prepared as in (Schmitt, Brown et al. 1990). Yeast cultures were grown to an OD_{600} of 1.0-1.5. Cells from a 10 ml culture were collected by centrifugation and resuspended in 400 ml AE buffer (50 mM NaOAc (pH 5.3), 10 mM EDTA (pH 8.0)). To this cell suspension, 40 ml 10% SDS and 450 µl of phenol (equilibrated with AE buffer) was added and the mixture were incubated at 65˚C for 4 minutes. The solution was frozen in dry-ice ethanol bath for ~30 seconds and then centrifuged for 6 minutes. The aqueous phase was extracted with phenol/chloroform (equilibrated with AE buffer) twice. 40 µl (1/10 volume) of 3 M NaOAc [pH 5.3], and 1 ml (2.5 volume) of ethanol were added to the aqueous phase to precipitate the RNA. The RNA was pelleted by centrifugation for 10 minutes, washed once with 1 ml of cold 70% ethanol, and resuspended in diethylpyrocarbonate (DEPC)-treated H_2O. RNA concentration was determined by measuring OD_{260} and calculated using the formula $1OD_{260}$=40 mg/ml.

Northern blotting

RNA was separated on a 1.5% agarose/formaldehyde gel (0.66M formaldehyde, 1XMOPS) in 1xMOPS buffer (0.04M morpholinopropanesulfonic acid [MOPS], 0.01 M sodium acetate, 1 mM EDTA, pH 7.2). The gel was soaked in DEPC-treated H_2O with shaking twice for 10 minutes each at room temperature to remove the formaldehyde. It was then soaked in 10XSSC with shaking for 10 minutes at room temperature. The RNA was transferred by capillary action to a

Genescreen*plus* membrane (DuPont NEN) in 10XSSC overnight. The membrane was baked at 80˚C for 2 hours and prehybridized in 50% formamide, 5XSSC, 5XDenhardt's solution, 1% SDS and 100 mg/ml denatured sheared salmon DNA at 50˚C for 4 hours.

Two types of probes were used, either antisense RNA or antisense DNA oligonucleotide. To make the antisense RNA probe, a template with a T7 promoter was made through PCR. Template for the Ppo.L1925 ribozyme probe was made with PrRS80 and PrRS79 using pJLI$_3\alpha$ as a template. Template for Tth.L1925 ribozyme probe was made with PrRS82 and PrRS49 using pRSTth-α as a template. Template for α fragment of β-galactosidase ribozyme probe was made with PrRS61 and PrRS81 using pRSTth-α as a template. 17S rRNA probe was made using PrRS111 and PrRS109 using yeast total genomic DNA as a template.

Synthesis of the antisense RNA probe was done with the MAXIscript *in vitro* transcription kit (Ambion Catalog number 1308-1326). Templates for the RNA probe were made by adding a T7 promoter at the 5' end of the complementary sequence of the desired probe through PCR, The full length Tth-α RNA probe template was made using primers PrRS83 and PrRS85 using pRSTth-α as a template. The α template was made using primers PrRS83 and PrRS62 using pRSTth-α as a template. The transcription mix consisted of, 5 ml of purified PCR product (purified with High Pure PCR Purification Kit, as above), 1 ml each of 10 mM ATP, CTP, and GTP, 40 μCi of labeled [α^{32}P]-UTP, 2 ml of 100 ml unlabeled UTP, 3 ml of nuclease free water, and 2 ml of T7 enzyme solution. The reaction was incubated at 37˚C for 1 hour. The reaction was passed through a Sephadex G-25 spin-column (Roche Diagnostics, catalog number 1814397) to remove excess label.

After purification, antisense RNA probes were added to the prehybridization solution. The membrane was hybridized for 16-24 hours at 50˚C and then

consecutively washed with 2XSSC/0.1% SDS for 15 minutes at room temperature, 0.5XSSC/0.1% SDS for 30 minutes at 60˚C, 1XSSC/0.1% SDS and 0.1XSSC/0.01% SDS each for 15 minutes each at room temperature.

For the DNA oligonucleotide probe, the probe was end-labeled. The end-labeling reaction consisted of 3 ml of the appropriate oligonucleotide (at 10 mM), 2 ml of 10x PNK buffer (New England Biolabs), 20 μCi [g^{32}P]-ATP, 1 μl PNK (10 U), and 13 μl DEPC-treated water. The reaction was incubated at 37˚C for 60 minutes, after which the PNK was heat-inactivated by incubating at 65˚C for 20 minutes. The reaction was passed through a Sephadex G-25 spin-column (Roche Diagnostics, catalog number NC9839095) to remove excess label. PrRS86 was used to probe the 5' end of α sequence. PrRS87 was used to probe the 3' end of Tth.L1925 sequence.

The probe was then added to the prehybridization solution. The membrane was hybridized for 16-24 hours at room temperature and then consecutively washed with 2XSSC/0.1% SDS for 15 minutes at room temperature and 0.5XSSC/0.1% SDS for 15 minutes at room temperature.

RNA size standards for the northern blots were also made with the MAXIscript *in vitro* transcription kit from Ambion (catalog number 1308-1326). PCR products were used as templates for transcription in a similar manner to the anti-sense RNA probes.

To strip oligo probes from northern blots, they were incubated with stripping solution (1% SDS, 0.1XSSC, 40 mM Tris, pH 7.5 mixed 50/50 (v/v) with formamide) twice at 50˚C for 60 minutes. To confirm that the probe had been removed from the blot, they were used to expose a phosophoimager plate overnight to check for removal of the probe. The blots were then re-probed as above.

Primer extension

Primer extension reactions were performed using the EndoFree RT kit (Ambion Catalog number 1740). Primer PrRS60 was used to prime the reaction. The corresponding sequence reaction was performed using the dsDNA cycle sequencing system (GibcoBRL Catalog number 18196), again using primer PrRS60. Both reactions were analyzed on an 8% polyacrylamide sequencing gel (Sequagel-8, National Diagnostics Catalog number EC-838). Gels were pre-run for 30 minuets at 65 W, max voltage, and run at 65 W, max voltage for 2 hours. After running, they were placed on filter paper and dried before exposing them to a phosphorimager plate, which was developed on a Storm phosphorimager (GE Healthcare).

Reverse transcription

Reverse transcription to detect circular intron RNA was performed using the RETROscript kit (Ambion, Catalog number AM1710). The RT reaction was performed using either PrRS71, which binds in the α region of both Tth-α and Ppo-α introns, or PrRS157 (for Tth.L1925-based introns) and PrRS158 (for Ppo.L1925-based introns), which bind in the P2 region. PCR amplification was performed using the RT primer and either PrRS103 (for Tth.L1925-based introns) or PrRS72 (for Ppo.L1925-based introns).

β-galactosidase assay

Assay for β-galactosidase activity was performed as described in (Ausubel, Brent et al. 1990). Yeast cultures were grown in appropriate medium to an OD_{600} of 1.0 to 1.5. Cells were collected by centrifugation and resuspended in an equal volume of Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol, pH 7.0). Aliquots of the cell-suspension were then transferred to another tube. Assays of low expressing cultures (~5-20 U) were assayed in 1 ml and 0.5 ml aliquots, while higher expressing cultures were assayed in 0.1 ml and 0.5 ml

aliquots. Z buffer was then added to bring the cell suspension to a final volume of 1 ml. Twenty ml of 0.1% SDS and forty ml of chloroform were added to the cell suspension and the suspension incubated at 30°C for 15 minutes. Two hundred ml of 4 mg/ml o-nitropphenyl-β-D-galactoside (ONPG) in 0.1 M KPO_4, pH7.0 buffer was added to start the reaction. At various time points, the reaction was stopped by adding 0.5 ml of 1 M Na_2CO_3. The reactions were then spun at 3000 rpm for 5 minutes, and then OD_{420} and OD_{550} of the supernatant were determined. β-gal activity units were calculated using the following equation: $U = 1000x(OD_{420}-(1.75xOD_{550}))/(TxVxOD_{600})$ where T = time of reaction in minutes, V = volume of cell suspension aliquot in ml, OD_{600} = cell density at the start of the assay.

Yeast protein extraction

Yeast protein extracts were made as in (Amberg, Burke et al. 2006). Yeast cells were grown in 6 ml of an appropriate medium to an OD_{600} of 1-3. Cells were collected and resuspended in 100 µl of H_2O, to which was added 100 µl of 0.2 M NaOH. The mixture was incubated at room temperature for 5 minutes. The yeast cells were pelleted by centrifugation for 3 minutes and the supernatant was removed. The pellet was resuspended in 50 µl 1x PAGE sample buffer (10% (by volume) glycerol, 62.5 mM Tris-HCL, pH 6.8, 2% SDS, 0.01 mg/ml bromophenol blue, 5% β-mercaptoethanol (BME)) and incubated in a boiling water bath for 3 minutes. Cell debris was pelleted and the supernatant was removed and transferred to a new tube. Protein concentration was determined using the NI Protein Assay Kit (G-Biosciences, Catalog number 786-005).

Western blotting

Yeast total protein extracts were run on 15% polyacrylamide gels on Might Small II eletrophoresis apparatus (Heofer, Catalog number SE250-10A-.75). Transfers were performed at 75 volts for 50 minutes in 1x transfer buffer (0.192 M glycine,

0.025 M Tris base, 20% methanol) onto an Immobilon-P membrane (Millipore). Blots were blocked in 1x TBST (150 mM NaCl, 20 mM Tris pH 8, 0.05% Tween 20) with 10% W/V with dry milk for at least 8 hours (usually over night) at 4˚C. Primary antibody (Monoclonal anti-HA antibody produced in mouse, Catalog number H 3663) was added to the blocking solution and incubated on a shaker for one hour at 4˚C. Blots washed five times for 10 minutes with 1x TBST with 1% W/V dry milk. The membrane was then immersed in 1x TBST with 1% W/V dry milk. The secondary antibody (Anti-mouse IgG from rabbit (whole molecule) conjugated to alkaline phosphatase) was added to this solution and incubated on a shaker for one hour at 4˚C. blots were washed five times for 10 minutes with 1x TBST with 1% W/V dry milk. Blots were removed from the wash and developed with 0.5 ml of ECF (GE Healthcare, Catalog number RPN5785) and visualized with a Storm phosphorimager (GE Healthcare). The HA-tagged standard used as a control for blots was a total *E. coli* lysate of cells overexpressing HA-tagged GST (Pierce catalog number 23613).

Chapter III

Chimeric *Tetrahymena* Introns Integrated in the rDNA of Yeast Express ORFs Inserted in the P1 Loop

Introduction

The group I self-splicing introns from *Physarum*, Ppo.L1925, and *Tetrahymena*, Tth.L1925, are very closely related. This might not be initially obvious, as Ppo.L1925 is nearly twice the size of Tth.L1925, but these introns share 70% sequence identity in their ribozyme regions and are inserted at homologous sites in the large ribosomal subunit (LSU) genes of their respective host organisms. The larger size of Ppo.L1925 is due to additional sequence encoding a homing endonuclease gene (HEG), I-PpoI, which Tth.L1925 lacks. It is likely that these two introns are descended from a common ancestor and that Ppo.L1925 acquired, or Tth.L1925 lost, a HEG after they diverged.

There is ample evidence that homing endonuclease genes (HEGs) can be gained or lost by group I introns. HEGs can transfer into group I introns through homing. There are two factors that influence whether the HEG will be able to successfully colonize the new host intron by homing (Figure 3.1). First, the HEG must not prevent the intron from splicing. At best, the disruption of intron splicing activity would render the intron's host gene non-functional (in the case of an unessential host gene); at worst, it would lead to the death of the host cell (in the case of an essential host gene). Second, the newly acquired HEG must be expressed. The advantage conferred upon the intron by the HEG, the ability to home into intron-lacking alleles, requires the protein the HEG encodes. If a HEG is not expressed, it confers no advantage to the intron; it would not be selected for, and would lose its ability to act as an endonuclease as mutations accumulate. Additionally, any future mutations that prevent the homing endonuclease protein from functioning would have

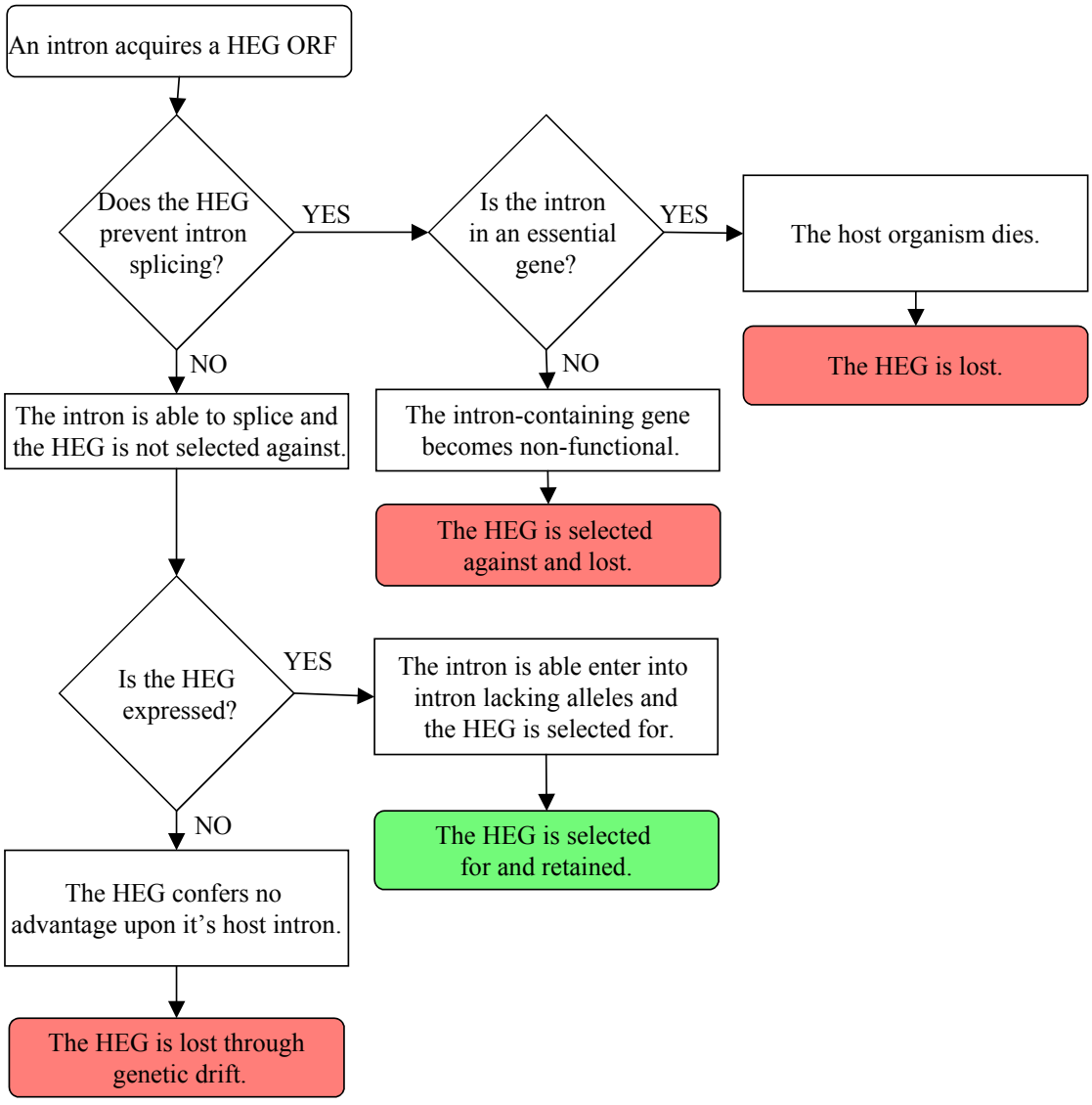


Figure 3.1: A flowchart detailing an intron's successful acquisition of a HEG. Red boxes indicate outcomes where the HEG is lost, while the green box indicates an outcome where the HEG ORF is retained by the intron.

a similar effect, and would lead to loss of HEG from the intron.

When functioning in the rDNA of yeast, Ppo.L1925 was shown to be able to express I-Ppo or two ORFs of similar size, the β-galactosidase α fragment and the yeast SOM1gene. Ppo.L1925 introns containing larger ORFs were unable to be integrated in yeast, so it is likely that beyond a certain size, ORFs disrupt accurate or efficient splicing (Lin and Vogt 2000).

The close relationship between Tth.L1925 and Ppo.L1925 suggest that the Tth.L1925 ribozyme could accommodate an inserted ORF and retain the ability to splice quickly and accurately. Indeed, the Tth.L1925 ribozyme is far more active than that of Ppo.L1925 (Zhang, Ramsay et al. 1995), so Tth.L1925 might be able to tolerate the large ORFs that Ppo.L1925 could not. Additionally, the model for intron acquisition of HEGs suggests that it takes few, if any, mutations to allow expression of an ORF once it is inserted in an intron. Thus, it seemed plausible that ORFs inserted into Tt.L1925 introns would not disrupt intron splicing and would be translated from the intron RNA. The experiments in this chapter were designed as an initial test of this hypothesis.

Results

Chimeric ORF-containing Tth.L1925 introns can trans-integrate into the yeast genome

The system I used to study these introns is to integrate the intron into all ~120 copies of ribosomal large subunit (LSU) of the budding yeast *Saccharomyces cerevisiae*. This system has several advantages over other *in vivo* systems, such as expressing the intron and surrounding RNA from a plasmid. In yeast, the ribosomal RNA genes (rDNA) are sequestered in the nucleolus, and nucleolus-specific factors, which may be important for intron splicing or expression of an intron-encoded ORF, might interact inefficiently, or not at all, with an intron transcribed from a plasmid. Additionally, since the intron is integrated into all ~120 copies of rDNA, it should be

easier to observe the protein products of weakly expressed intron-encoded ORFs than it would be from the ~20 copies of a yeast expression plasmid. The one major disadvantage of this system is that it requires the intron to be able to self-splice. Integration of a non-splicing intron would prevent the formation of ribosomal RNA and thus prevent the yeast cell from growing.

Introns are introduced into the yeast rDNA through the process of trans-integration (Figure 1.6). The intron of interest, flanked by yeast rDNA sequences, is constructed in a yeast shuttle vector. This shuttle vector is co-transformed with pCPIPpo, a plasmid which encodes the homing endonuclease I-PpoI under Gal 1,10 promoter control, into the desired yeast strain. Inducing expression from the plasmid causes a reaction similar to homing; the endonuclease cleaves the yeast chromosomal DNA at the intron insertion sites. The rDNA exon sequence flanking the intron allows the intron-containing plasmid to act as template for double strand break repair (DSBR). Since the intron sequence interrupts the homing endonuclease target site, cut rDNA copies that use the intron-containing as sequence as a template for DSBR are resistant to cleavage by the homing endonuclease, while cut rDNA copies that use intron-lacking rDNA copies as a template for DSBR will be re-cleaved by the homing endonuclease. This results in the intron sequence being copied into all the rDNA copies of the yeast cell. However, there is a second way for the rDNA to become immune to cleavage by the homing endonuclease. Sometimes, during the repair of an rDNA repeat using a intron-lacking template, mutations that make the repaired allele immune to homing endonuclease cleavage are introduced to the repaired allele. If one of these mutations occurs, the mutant rDNA can be used as a template for DSBR, instead of the intron, and thus could spread rapidly to all the rDNA copies. Potential trans-integrants can be screened for the presence of the intron or for the homing endonuclease resistance mutation through PCR amplification over the intron insertion

site. Since the intron-donor plasmid contains some rDNA sequence, primers complementary to regions outside of the donor plasmid rDNA sequence are used to avoid amplifying sequences from any residual intron-donor plasmid. Normally, potential trans-integrants contain only intron-integrated rDNA copies or only endonuclease-resistant rDNA copies, but not mixtures of the two. In rare cases, a mixture of intron-inserted and endonuclease-resistant rDNA has been found. This phenotype has only been seen with introns that splice poorly, and presumably, the endonuclease-resistant rDNA copies are necessary to provide enough rRNA for yeast cell to continue growing.

The Tth-ClaI intron served as a starting point and a useful tool for my experiments. This intron is based on Tth.L1925, but four nucleotides in the P1 loop were mutated by Jue Lin to encode a ClaI restriction site (Figure 3.2A, B). The site of the ClaI site is homologous to the site of the I-PpoI ORF in Ppo.L1925 (Figure 3.2 C, D). Using this ClaI site, I could then insert ORFs into the P1 loop to create chimeric Tth.L1925 introns. Tth-ClaI can be trans-integrated into yeast (Figure 3.4), which indicates that the four-nucleotide change does not disrupt splicing.

I chose four ORFs to insert into the ClaI site of Tth-ClaI: the α fragment of β-galactosidase, Neo, *HIS3*, and I-PpoI (Figure 3.3). The β-galactosidase α fragment is a small ORF, ~300 base pairs in length, and can replace the I-PpoI ORF in Ppo.L1925 without disrupting splicing. The bacterial Neo and yeast *HIS3* are two larger ORFs, both between 700-800 base pairs in length. Since chimeric *Physarum* introns encoding Neo and *HIS3* are unable to integrate into the yeast genome, these ORFs presumably disrupt splicing in that context. Lastly, I-PpoI is the HEG from Ppo.L1925. Successful I-Ppo expression from Tth.L1925 would demonstrate that a homing endonuclease can be introduced into an intron sequence and be immediately expressed.

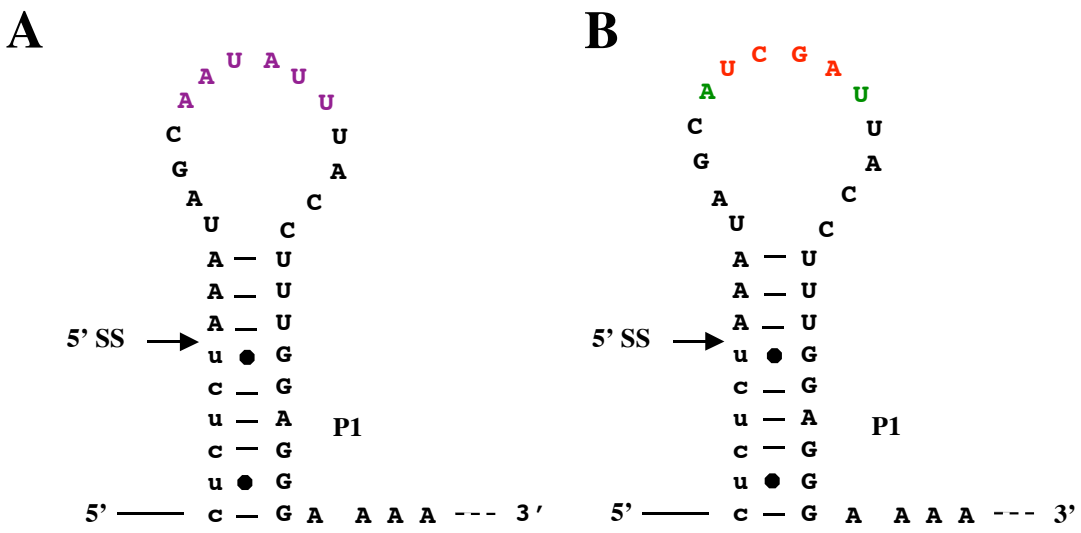


Figure 3.2: The P1 stem loops of Tth.L1925, Tth-ClaI, a chimeric Tth.L1925 intron based on Tth-ClaI, and Ppo.L1925. A. Tth.L1925. B. Tth-ClaI. Purple nucleotides indicate the wild-type Tth.L1925 sequence in the region mutated to make Tth-ClaI. Red nucleotides indicate mutations necessary to add the ClaI site to the P1 loop. Green nucleotides are part of the ClaI site but are unchanged from the wild type Tth.L1925 sequence. The 5' exon sequence is in lowercase.

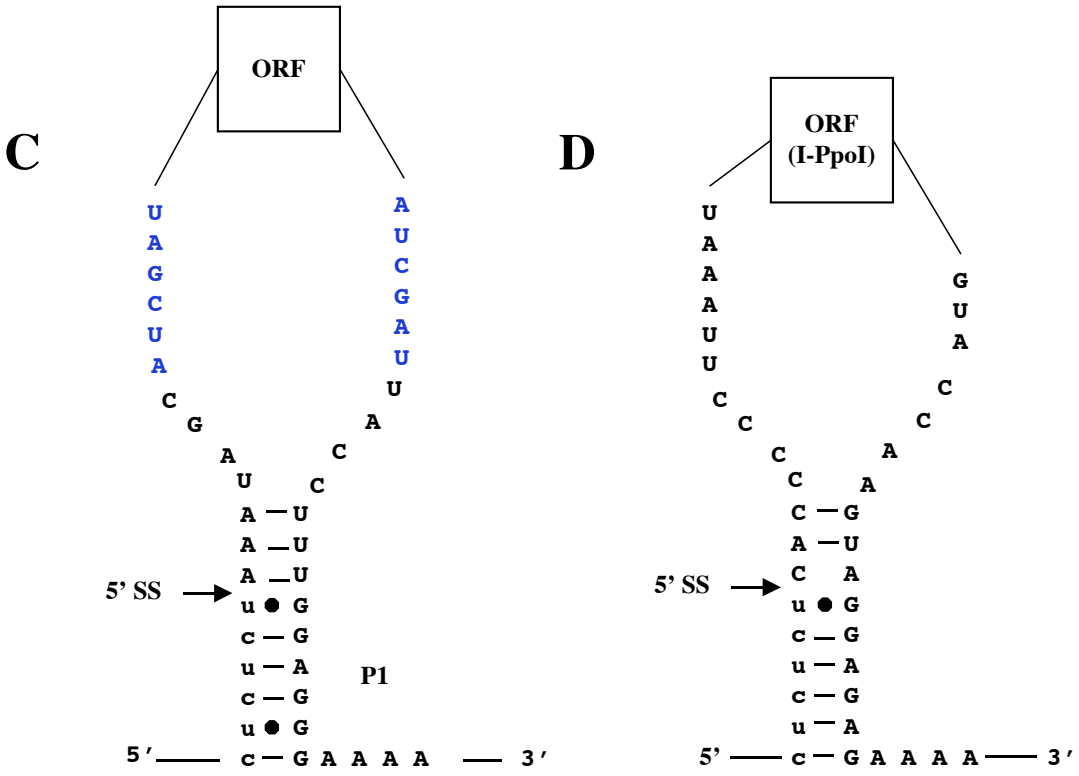
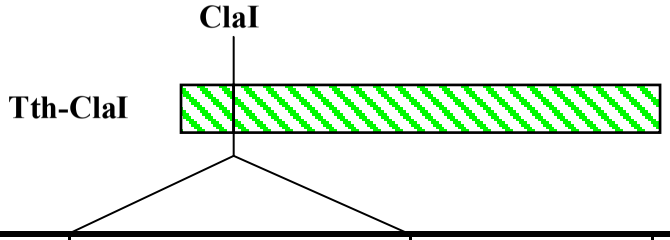


Figure 3.2 (Continued) . The P1 stem loops of: C. A chimeric Tth.L1925 intron based on Tth-ClaI. D. Ppo.L1925. Blue nucleotides indicate the two ClaI sites in the chimeric Tth-ClaI intron. The 5' exon sequence is in lowercase.



Intron Name	Inserted ORF	ORF length	Chimeric Ppo.L1925 with this ORF splices?	Is this ORF expressed from chimeric Pp.L1925?
Tth-α	β-galactosidase α fragment ORF	297 bp	Yes	Yes
Tth-Neo	Neo ORF	795 bp	No	No
Tth-His3HA	His3 ORF with 2 HA tags at the 3' end	717 bp	N/A (No for His3 without HA tags)	N/A (No for His3 without HA tags)
Tth-IPpoHA	I-PpoI ORF with 2 HA tags at the 3' end	549 bp	N/A (Yes for I-PpoI without HA tags and 3'UTR sequence)	N/A (Yes for I-PpoI without HA tags and 3'UTR sequence)

Figure 3.3. The four chimeric Tth.L1925 introns constructed in this chapter. The category "Chimeric Ppo.L1925 with this ORF splices?" indicates if a chimeric Ppo.L1925 intron with the indicated ORF could be trans-integrated into *S. cerevisiae*.

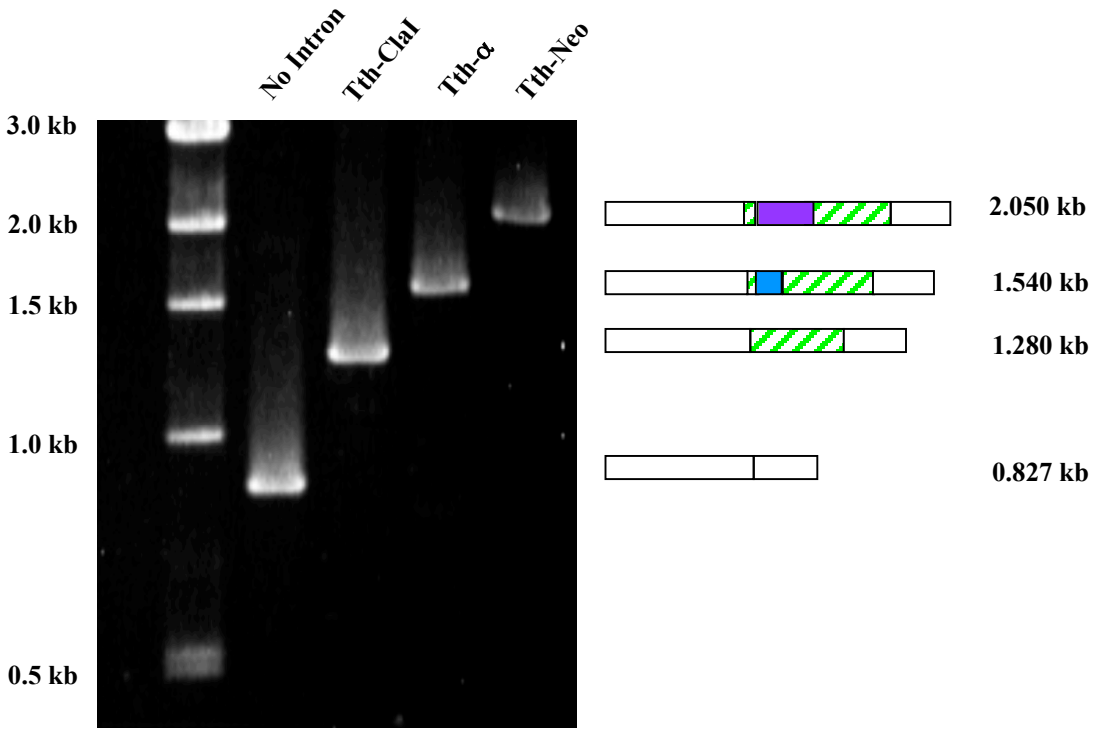


Figure 3.4: Trans-integration of chimeric Tth.L1925 introns. Agarose gel electrophoresis of PCR products from yeast DNA extracts. Lane 1 1Kb Ladder (New England Biolabs). INVSc2-the PCR product of non-intron integrated yeast strain InvSCII. Tth-ClaI: PCR product from total genomic DNA of INVSc2 yeast with integrated Tth-ClaI. Tth-a: PCR product from NOY505GalWT yeast with integrated Tth-α. Tth-Neo: PRC product from INVSc2 yeast with integrated Tth-Neo.

Tth.L1925 is able to accommodate and express a small inserted ORF

The first ORF I inserted into Tth-ClaI was the α-fragment of β-galactosidase. Since this ORF had already been shown not to interfere with splicing of, and to be expressed from, the α-containing *Physarum* intron, Ppo-α, I assumed it would have the highest probability of being expressed from Tth.L1925. I cloned the β-galactosidase α ORF into the ClaI site of Tth-ClaI to create the intron Tth-α. Because it was thought that the nucleotides near the 5' and 3' ends of the ORF could be important in splicing, Ppo-α had been designed to conserve the first and last fifteen nucleotides of I-*Ppo* I sequence. The α ORF was inserted into a ClaI site engineered between the 5' and 3' 15-nucleotides, resulting in an α protein with seven additional amino acids at its N and C termini (five I-Ppo I amino acids and two from the 6 nucleotide ClaI site). In order to make Tth-α as directly comparable to Ppo-α as possible, I designed the intron to encode a similar ORF. The Tth-α ORF keeps the first and last five amino acid residues of the I-Ppo I sequence. Since I was cloning the α ORF into the external ClaI site of Tth-ClaI, I mutated the internal α ClaI sites in Ppo-α ORF to prevent ClaI cleavage and to conserve the two amino acids encoded.

The β-galactosidase α fragment is a small piece of the larger β-galactosidase protein. The α fragment has no activity of its own, but when expressed *in trans* with the remainder of the β-galactosidase protein, the Ω fragment, the two proteins associate. The resulting α-Ω complex has full β-galactosidase activity. This phenomenon, known as α-complementation, was demonstrated in yeast for the first time during analysis of Ppo-α (Lin and Vogt 2000). The yeast strain NOY505GalΩT is a useful tool for measuring the amount of α protein being produced by an integrated intron, as this strain has the Ω-fragment gene integrated into the trp1 locus under Gal 1,10 promoter control (Figure 3.5A). NOY505GalΩT is also the same yeast strain that was used to analyze expression from Ppo-α (Lin and Vogt 2000). In this system, the

α-fragment expressed from the intron complements with the Ω-fragment expressed from the chromosomal locus (Figure 3.5B).

Tth-α is able to splice efficiently, as the majority of potential trans-integrants screened contained integrated introns (Figure 3.4). The α ORF is expressed from the integrated intron, as these yeast cells exhibited a light bluish color on media containing X-gal (data not shown). ONPG assays on liquid cultures indicated that the β-galactosidase activity of Tth-α was 10 U, slightly less than the 15 U of Ppo-α (Figure 3.5D). Assuming that the specific activity for α complementing β-galactosidase is comparable to that of the whole molecule, (300,000 U/mg (Steers, Cuatrecasas et al. 1971)), I estimate that α protein is 0.05% of yeast total protein. It is unlikely that the amount of α protein is being underestimated, for the following reason. Crude extracts from NOY505GalΩT transformed with a multi-copy plasmid expressing the α-fragment, which presumably produces α in excess of the Ω fragment produced from a single copy gene, exhibited roughly 500 Units of β-galactosidase activity. This is roughly 25% of the activity seen from the intact β-galactosidase protein expressed from the same plasmid (Lin and Vogt 2000), which greatly exceeds the activity seen from Tth-α, suggesting that the Ω fragment is in excess of the α fragment in the Tth-α-integrated yeast.

Tth.L1925 is able to accommodate and express large inserted ORFs

It is clear that Tth.L1925 can accept a small ORF, but is it able to accept and express larger ORFs? To address this question, I chose to insert two ORFs, Neo and *HIS3*, both of which had been shown earlier to disrupt splicing of Ppo.L1925 (Lin and Vogt 2000). Both of these reporter genes give a qualitative phenotype: Neo confers resistance to G418 and *HIS3* allows growth on -His media. I decided to add an hemagglutinin (HA) epitope tag to the end of the *HIS3* ORF to allow for a quantitative measurement of the amount of *HIS3* protein being produced, through western blotting

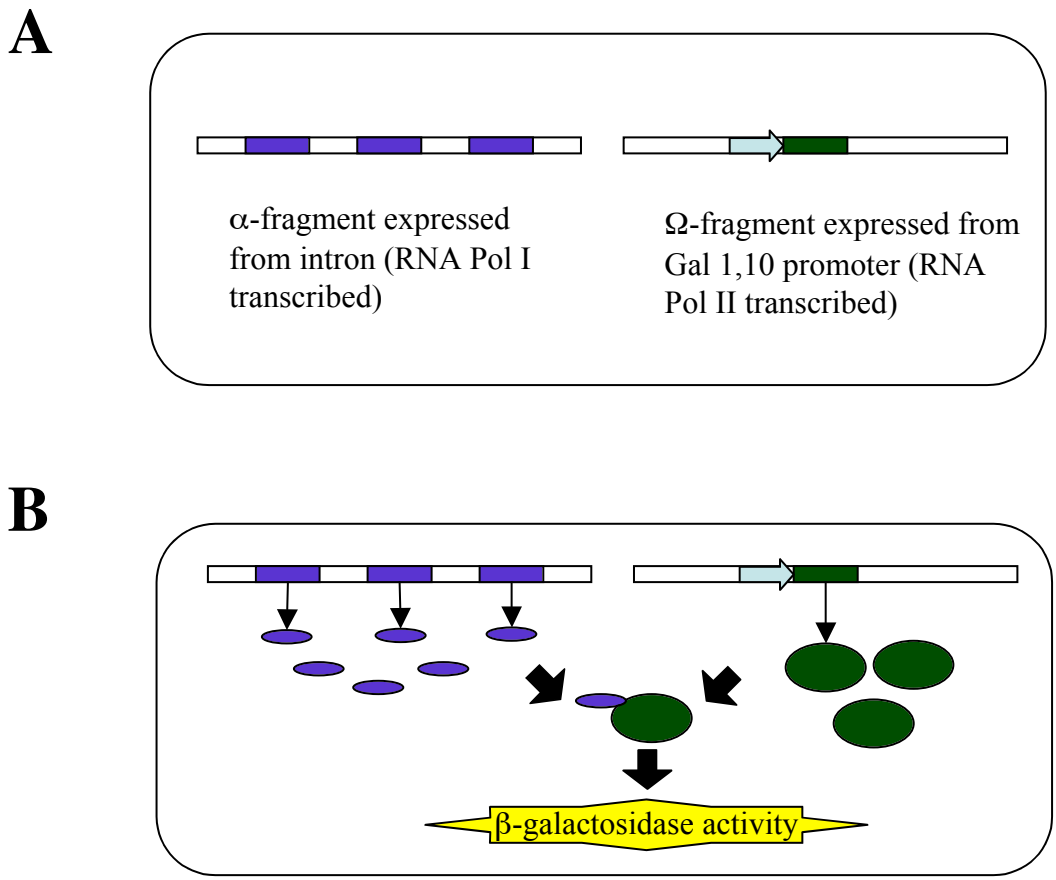
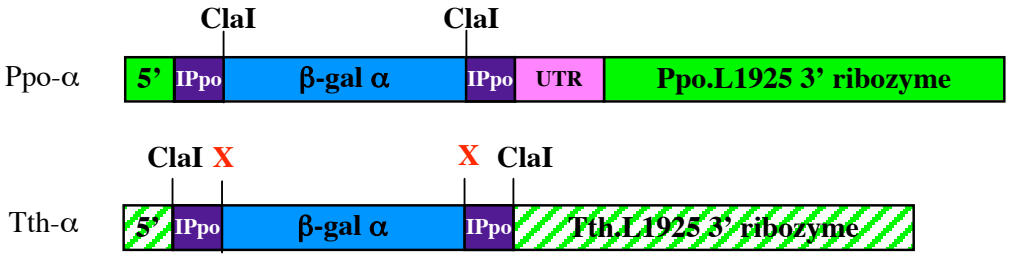


Figure 3.5: Tth-α. A. The NOY505GalΩT system. The α fragment is encoded in introns integrated in the ~120 copies of rDNA, while the Ω fragment is expressed from a single copy under Gal 1,10 promoter control. B. The α fragment expressed from the introns complements with the Ω fragment expressed from the single locus to produce β-galactosidase activity.

C



D

Intron	β-galactosidase activity (U)
No intron	0.4
Integrated Tth-ClaI	0.2
Integrated Ppo-α	15
Integrated Tth-α	10
Plasmid pRSTth-α	0.4

Figure 3.5 (Continued): C. The Ppo-α and Tth-α introns. The 15 5' and 3' I-Ppo nucleotides are indicated in purple. The mutated ClaI sites in Tth-α are indicated by the X's. D. β-galactosidase activity of introns integrated into yeast strain NOY505GalΩT. Plasmid pRSTth-α indicates NOY505GalΩT transformed with the plasmid used for trans-integrating Tth-α.

of yeast protein extracts. The HA tag is a nine-amino acid sequences (YPYDVPDYA) from the human influenza virus hemagglutinin protein that is commonly used as an epitope tag (Field, Nikawa et al. 1988). I inserted two copies of the HA sequence interrupted by a single glycine residue in order to increase the number of binding sites for anti-HA antibodies (Figure 3.7B). Because of the possibility that the sequence encoding the HA tag itself might disrupt intron splicing or somehow inhibit translation, I decided to only add the HA tag to the *HIS3* ORF and excluded it from the Neo ORF. Additionally, to determine if the five upstream and downstream I-PpoI amino acids used in Tth-α are required for splicing or translation, I chose to omit them from the Neo and His3-HA ORFs.

I cloned these two ORFs into the ClaI site of Tth-ClaI to create introns Tth-Neo (Figure 3.6) and Tth-His3HA (Figure 3.7) and trans-integrated them into *S. cerevisiae*. PCR analysis of isolated colonies potentially carrying Tth-Neo (Figure 3.4) and Tth-His3HA (data not shown) introns showed that some of them contain integrated introns. Thus, the Tth.L1925 ribozyme is able to splice with large inserted ORFs that disrupt splicing when inserted in its cousin intron Ppo.L1925. Also, the HA tag sequence added to His3 ORF, and the lack of the sequence encoding the 5' and 3' I-PpoI amino acids from both ORFs, did not appear to significantly disrupt splicing.

If the Neo protein is being expressed from Tth-Neo, yeast cells with integrated Tth-Neo should be resistant to G418. Yeast with integrated Tth-Neo grew on media containing G418, while the Tth-Neo-lacking parent strain did not (Figure 3.6), indicating that Neo expression from Tth-Neo is indeed occurring.

Due to the added HA-tag, expression of the His3HA can be tracked both qualitatively, by plating the yeast on –His selective media, and quantitatively, through western blotting of protein extracts and comparison with a known standard. Yeast with integrated Tth-His3HA introns were able to grow on –His media, indicating that

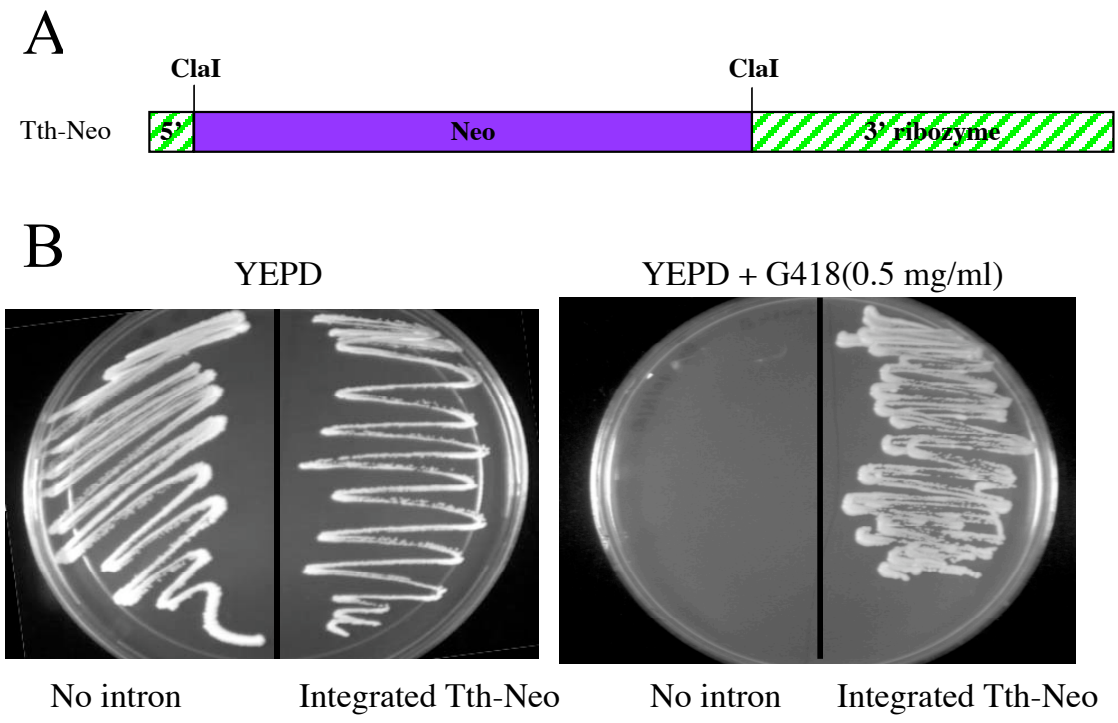


Figure 3.6: Tth-Neo. A. Schematic representation of the Tth-Neo intron. B. Yeast with no integrated intron or with integrated Tth-Neo are plated on YEPD (left) and YEPD + 0.5 mg/ml G418.

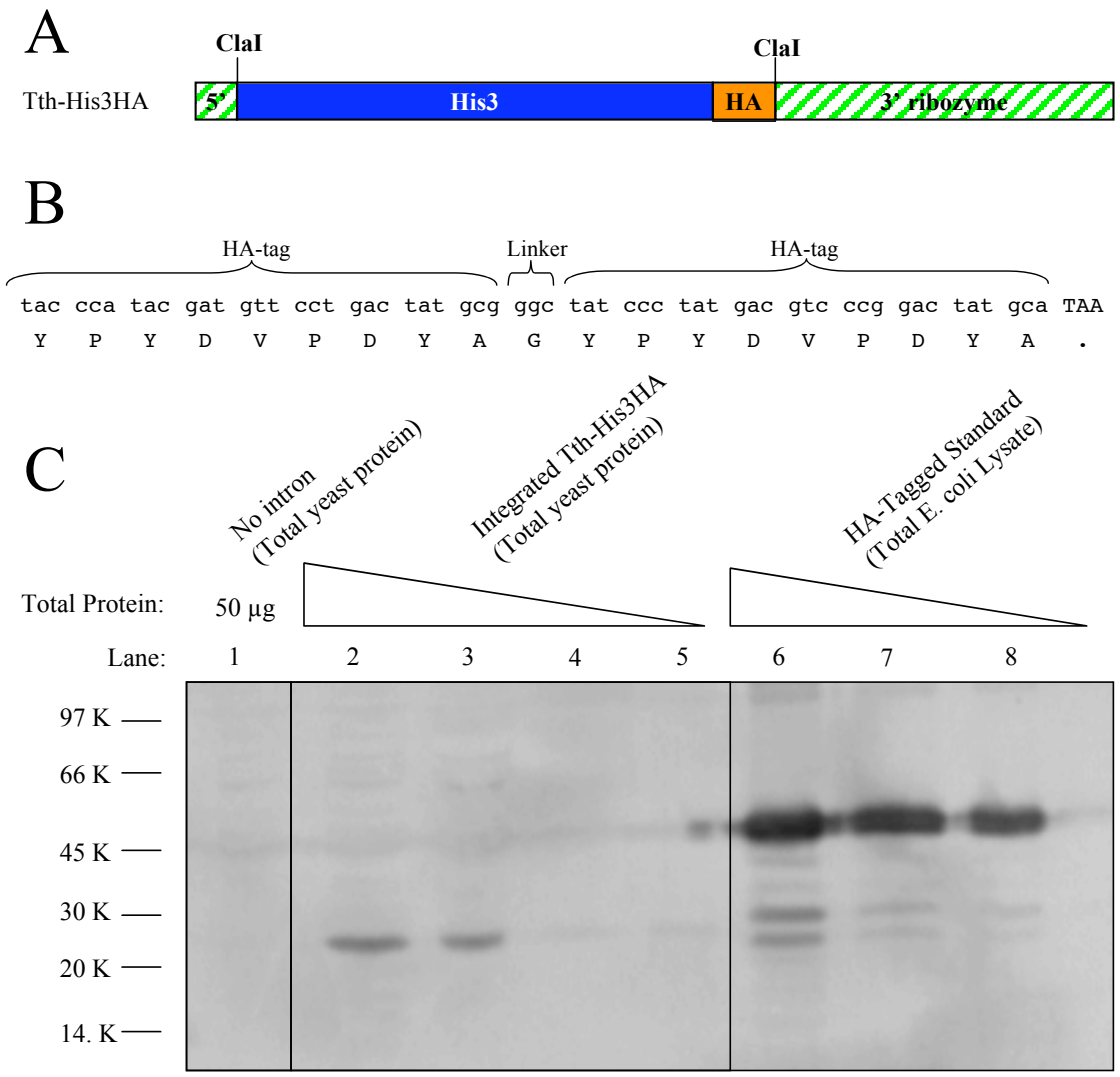


Figure 3.7: Tth-His3HA. A. Schematic representation of the Tth-His3HA intron. B. Sequence of the double HA tag added to the 3' end of the His3 ORF. C. Western blot of yeast extracts probed with an anti-HA antibody. Lane 1, 50 µg of total Noy505GalΩT protein. Lane 2-5, 50 µg, 25 µg, 6.25 µg, and 2.5 µg of Noy505GalWT with integrated Tth-His3HA total protein. Lanes 6-8, 0.25 µg, 0.05 µg, and 0.0125 µg of total lysate of E. coli over expressing a HA tagged standard.

the His3HA protein was being produced by the cells. The protein could be visualized by western blot (Figure 3.7).

To estimate the amount of His3-HA produced, I used the total lysate of an *E. coli* strain which was overexpressing a singly-tagged HA standard. To estimate the percentage of the lysate that the HA-tagged standard comprised, I analyzed Coomasie stained polyacrylamide gels using Adobie Photoshop and determined that the standard made up approximately 15% of the lysate. I then compared the intensity of the standard band in western blots to that of the His3HA using the ImageQuant. From this comparison, I estimate that the His3HA protein comprises ~0.01% of the yeast total protein, an amount similar to that of β-galactosidase a fragment produced by Tth-α.

Insertion of the I-PpoI ORF disrupts splicing of chimeric-Tth.L1925 introns

The fourth ORF I examined was the I-PpoI HEG. Inserting this ORF into the Tth-ClaI intron is interesting from an evolutionary perspective, as it is likely that the HEG that was either lost or gained by an ancestral Tth.L1925 and Ppo.L1925 intron after their divergence from a common ancestral intron. The 489 nucleotide I-PpoI ORF is larger than the 255 α ORF it and in its native context, obviously does not disrupt the splicing of Ppo.L1925.

To make the chimeric intron Tth-IPpoHA, I decided to add the double HA-tag to the 3' end of the I-PpoI ORF, before insertion into Tth-ClaI (Figure 3.8). The amount of I-PpoI protein produced in principle could be quantitated through an endonuclease activity assay, where substrate DNA containing the 15 nucleotide I-PpoI cleavage site is incubated with a protein extract, or by western blotting. Crystal structures of I-PpoI show that the C-terminus, to which the HA tag is appended, is on the surface of the protein and is involved in making dimerization contacts, and is not involved in DNA binding or catalytic sites, so it seemed likely that the C-terminus

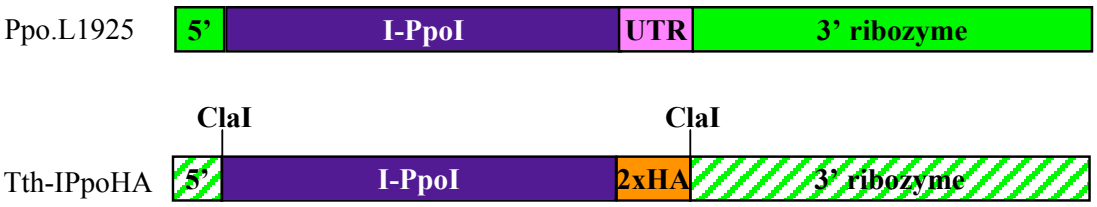


Figure 3.8: Tth-IPpoHA. A. Schematic representation of the Ppo.L1925 and Tth-IPpoHA introns. The HA sequence used in Tth-IPpoHA is identical to that used in Tth-His3HA (Figure 3.7).

would be able to accommodate the extra HA-tag amino acids (Flick, Jurica et al. 1998).

My attempts to trans-integrate Tth-IPpoHA were unsuccessful. All of the 12 potential Tth-IPpoHA integrants screened from two independent trans-integration reactions contained I-PpoI resistant rDNA but no integrated introns (data not shown). This suggests that the Tth-IPpoHA intron is unable to splice efficiently in *S. cerevisiae*. It is unclear how the ORF is disrupting splicing, although is unlikely that this disruption was caused by ORF length, as introns with the longer His3-HA and Neo ORFs were able to splice.

In summary, three out of the four tested chimeric Tth.L1925 introns, Tth-α, Tth-Neo, and Tth-His3HA were able to splice, despite the inserted ORF. This result is in contrast to previous observations that Neo and His3 prevented splicing in chimeric Ppo.L1925 introns. The I-PpoI ORF prevented splicing in the context of Tth-IPpoHA Tth-α Tth-Neo, and Tth-His3HA produced functional proteins, with the latter two being at comparable levels, ~0.01% of total yeast protein. Protein production obviously was very inefficient, much lower per gene copy number than from a RNA Pol II-transcribed gene.

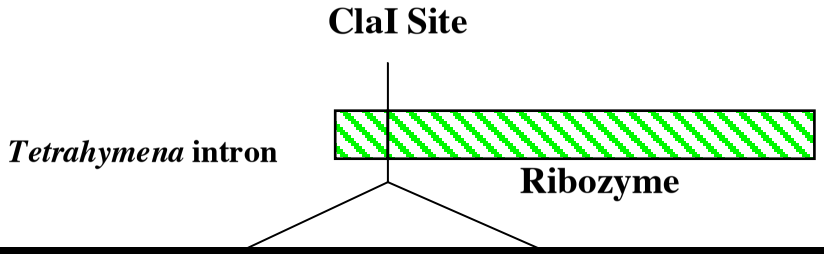
Discussion

In order to be successful in its parasitic lifestyle, a group I intron must be able to remove itself from the primary transcript of its host gene. The ability of an intron to splice is affected by both the catalytic activity of the ribozyme and the rate that the freshly-synthesized ribozyme RNA folds into the active conformation. The intron-folding rate is affected by surrounding exon sequences, the presence of inserted sequences, and the presence of splicing factors (protein factors that promote, or act as a scaffold for, intron folding) (Woodson and Cech 1991; Rocheleau and Woodson 1995; Koduvayur and Woodson 2004). Misfolded intron RNAs do not appear to get a

second chance to fold correctly and are instead sequestered and degraded in yeast (Jackson, Koduvayur et al. 2006). Because of this degradation of misfolded RNA, a slowly-folding intron that forms a highly active ribozyme may appear to be inactive *in vivo*.

In order to be accepted into an intron sequence, an inserted ORF cannot interfere too severely with either the folding pathway of the ribozyme RNA or the splicing reaction of the ribozyme. The inserted ORF sequence would disrupt splicing by blocking important catalytic ribozyme sites, such as the exogenous G binding site. It could also promote mistargeting of the 5' or 3' splice sites, which would result in excision of exon sequence or inclusion of some intron sequence in the spliced exon RNA, disrupting structures formed by exon RNA and resulting in frame-shift mutations in protein coding RNAs. ORF sequences could disrupt folding by pairing with ribozyme sequences before they can form their proper pairing, or the ORF could be so long as to prevent ribozyme elements interrupted by the ORF from pairing before the unfolded pre-rRNA is degraded.

Tth.L1925 is able to accommodate the small β-galactosidase α ORF as well as the larger Neo and His3HA ORFs and retain the ability to function *in vivo* (Figure 3.9). The smaller α ORF can be inserted in both Tth.L1925 and Ppo.L1925 introns without disrupting splicing. The larger Neo and His3HA ORFs disrupt the splicing of Ppo.L1925 but not that of Tth.L1925. I interpret this difference to be due to the robustness of the Tth.L1925 ribozyme, which would allow Tth.L1925, even if splicing were slowed by and ORF, to splice at a fast enough rate to produce functional rRNA. These results support the idea that a HEG ORF could enter into an intron sequence without disrupting intron function, preventing selection against the newly ORF-containing intron.



Intron Name	Inserted ORF	ORF length	Chimeric Tth.L1925 integrates?	ORF expressed from Chimeric Tth.L1925?	ORF expressed as % of total yeast protein.
Tth-α	β-galactosidase α fragment ORF	297 bp	Yes	Yes	0.05%
Tth-Neo	Neo ORF	795 bp	Yes	Yes	N/M
Tth-His3HA	His3 ORF with 2 HA tags at the 3' end	717 bp	Yes	Yes	0.01%
Tth-IPpoHA	I-PpoI ORF with 2 HA tags at the 3' end	549 bp	No	N/A	N/A

Figure 3.9: Summary of the chimeric introns studied in Chapter III. N/M indicates not measured, while N/A indicates that no measurements were possible, as the associated intron could not be trans-integrated.

The one tested chimeric Tth.L1925 intron that is unable to trans-integrate into the yeast genome is Tth-IPpoHA, implying that the inserted ORF disrupts folding or splicing, or that the protein or RNA is otherwise toxic to the yeast. It is possible that I randomly selected twelve I-PpoI resistant mutants out of a larger pool of yeast that contained both I-Ppo resistant mutants and integrated introns. However, this possibility seems unlikely, as the majority of potential integrants of other chimeric Tth.L1925 introns screened contained integrated introns while a minority contained resistant mutants.

In Ppo.L1925, the I-PpoI ORF is followed by a 53-nucleotide sequence dubbed the 3' untranslated region (3'UTR), which is not part of the Ppo.L1925 ribozyme structure and is not present in Tth.L1925. Deleting nucleotides 41-50 of the 3' UTR prevents any trans-integration of Ppo.L1925 and full trans-integrating of Ppo-α, so this UTR sequence is important for splicing (Lin and Vogt 2000). Interestingly, the actual sequence of nucleotides 41-50 does not seem to affect splicing, as introns with a wide variety of different sequences in this region are able to trans-integrate. It could be that the 3'UTR sequence is required for the splicing of an I-PpoI-containing intron, perhaps acting as a spacer and preventing the I-PpoI sequence from interacting with ribozyme sequence. Doubt is cast on this possibility, though, by the fact that the 3'UTR sequence is required for splicing of Ppo.L1925-based introns that do not contain the I-PpoI sequence. For example, the 3'UTR is required for Ppo-α splicing, but is not required for the splicing of Tth-α.

Another possibility is that Tth-IPpoHA is actually able to trans-integrate and splice and the I-Ppo I-HA protein is somehow toxic to yeast cells. The I-PpoI protein is toxic when expressed in yeast cells that do not contain I-PpoI resistance mutations in their rDNA, and, if the I-Ppo I-HA protein is produced high levels, it might cut the host cell DNA at a rate that the DSBR machinery is unable fully correct. However,

this possibility is rendered unlikely, as during the trans-integration process, I-PpoI is expressed at high levels from the Gal 1,10 promoter from the multi-copy pCPIPpo plasmid, yet some yeast survive. In addition, this scenario is incongruent with the levels of expression seen from Tth-α and Tth-His3HA, both of which are much lower than that of a Gal 1,10 promoter.

A third possibility is that the HA tag added at the C-terminal end of I-PpoI interfered with DNA recognition and perhaps reduced the specificity of the I-PpoHA enzyme. This could cause the protein to cut the yeast genome at sites other than the I-PpoI recognition site, which would be highly toxic to the yeast cells. If this is the case, it is unlikely that yeast with integrated Tth-IPpoHA would survive. A crystal structure of I-PpoI bound to DNA shows, though, that the C-terminus of the I-PpoI protein is involved in dimerization of I-PpoI proteins and is relatively far away from the active and DNA binding sites of the protein (Flick, Jurica et al. 1998). If the HA tag interfered with I-PpoHA dimerization, it would lower I-PpoI activity and actually make the protein less toxic.

Expression of chimeric Tth.L1925 encoded ORFs

Based on the results in this chapter, it appears that no significant changes in the intron sequence is required for expression of most ORFs inserted in the P1 loop of Tth.L1925. The three chimeric Tth.L1925 introns that are able to integrate into all of the rDNA copies of yeast are expressed.

Estimates of the amount of protein produced from Tth-α and Tth-His3HA indicate that the introns are producing roughly similar amounts of protein, 0.05% and 0.01%, respectively, of total yeast protein. These levels are consistent with the amount of protein produced by both Ppo.L1925 and Ppo-α in yeast, where I-PpoI is 0.04% and α is 0.07% of yeast total protein, respectively (Lin and Vogt 1998; Lin and Vogt 2000). Protein is expressed from either Tth.L1925 or Ppo.L1925 much less

efficiently, roughly 1000-fold less per DNA copy, than from a highly active RNA Pol II promoter such as the Gal 1,10 promoter. Interestingly, protein expression from Ppo.L1925 in its native genome of *Physarum* is even lower, with I-PpoI representing $1.3x10^{-4}$% of *Physarum* total protein (Lin and Vogt 1998).

Chapter IV

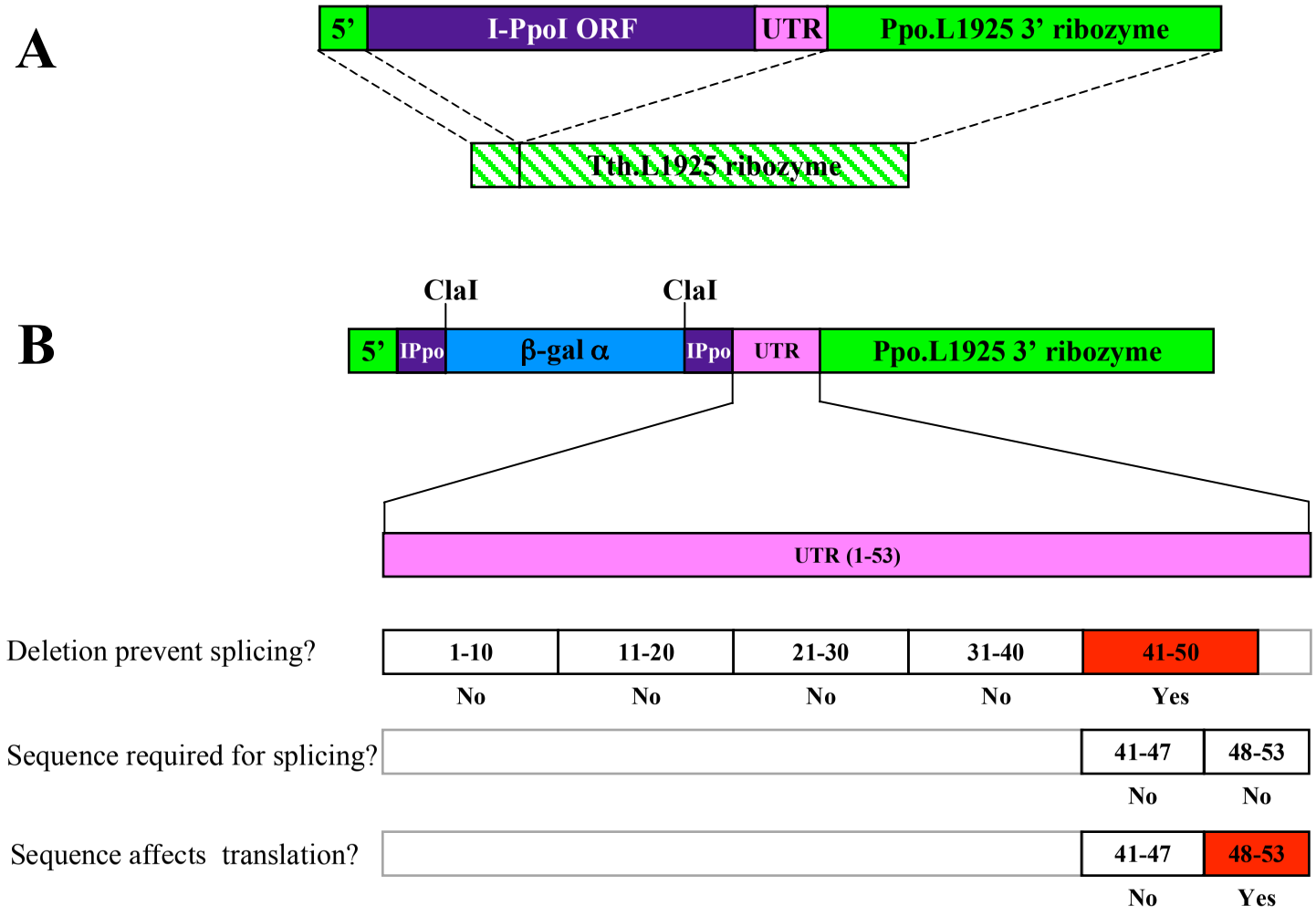
The Role of Intron Sequence Elements in Expression from Chimeric *Tetrahymena* Introns.

Introduction

The most obvious difference between the Ppo.L1925 and Tth.L1925 introns is the large I-PpoI ORF in Ppo.L1925. In addition, there are also other major sequence differences between these two introns (Figure 4.1 A). One of these regions, a 53-nucleotide sequence called the 3' untranslated region (3'UTR), lies between the end of the I-PpoI ORF and the Ppo.L1925 ribozyme and is entirely absent in the Tth.L1925 intron. The short 5' end sequences of the introns are also very different, having identical bases at only 3 of the first 11 nucleotides, compared with the 70% sequence identity in the core ribozyme.

The role of the 53-nucleotide 3'UTR sequence has been investigated in Ppo.L1925, using both Ppo.L1925 and the chimeric Ppo-α (Lin and Vogt 1998; Lin and Vogt 2000). After the Ppo.L1925 intron is spliced, a processing cleavage separates the 3'UTR and the upstream ORF from the Ppo.L1925 ribozyme RNA (Figure 1.13). Initially, it was hypothesized that the 5' processed RNA functioned as the message for I-PpoI translation and that the 3'UTR stabilized the RNA or somehow promoted translation. Experiments showed that instead, the full-length intron RNA apparently serves as I-PpoI message (Lin and Vogt 1998). Thus, it appears that the functional 3' untranslated region of the Ppo.L1925-encoded ORF consists of both the 53-nucleotide region and the entire Ppo.L1925 ribozyme. This means that referring to the 53-nucleotide sequence as the 3'UTR is a misnomer, even though the term is still used for this region.

Figure 4.1: Features of Tth.L1925 and Ppo.L1925. A. Schematic representation of sequences shared between Ppo.L1925 (top) and Tth.L1925. The Ppo.L1925 ribozyme region is interrupted by the I-PpoI ORF and the 53 nucleotide 3'UTR sequence, while Tth.L1925 has only a ribozyme region. B. Summary of previous experiments examining the role of the 3'UTR in splicing and expression. Numbers indicate 3'UTR nucleotide positions (from 1 to 53).


A
5'
I-PpoI ORF
UTR
Ppo.L1925 3' ribozyme
T1h.L1925 ribozyme
B
ClaI
ClaI
5'
IPpo
β-gal α
IPpo
UTR
Ppo.L1925 3' ribozyme
UTR (1-53)
Deletion prevent splicing?
1-10
11-20
21-30
31-40
41-50
No
No
No
No
Yes
Sequence required for splicing?
41-47
48-53
No
No
Sequence affects translation?
41-47
48-53
No
Yes

The function of the 3'UTR is mysterious, but it appears to play a role in both intron splicing and ORF expression (Figure 4.1 B) (Lin and Vogt 2000). By sequentially deleting 10 nucleotide sections of the 3'UTR, Jue Lin found that nucleotides 1-10, 11-20, 21-30 and 31-40 have no significant effect on splicing, as inferred from the ability of the chimeric α intron to integrate into yeast rDNA. Deleting 3'UTR nucleotides 41-50, however, appeared to greatly reduce intron splicing. A possible explanation for this finding was suggested by the observation that nucleotides 46-53 are complementary to, and might be base-pairing with, an eight nucleotide stretch in the I-PpoI ORF. However, this hypothesis was disproven, as changing the I-PpoI sequence so that it is no long complementary to the 3'UTR sequence had no effect on splicing (Lin and Vogt 2000). Additionally, experiments that created pools of introns with random nucleotides at positions 41-47 and 48-53 showed that a wide variety of sequences at these positions did not disrupt splicing. This result strongly suggests that the sequence of the nucleotides is not important for splicing. On the other hand, the sequence of the last six nucleotides (48-53) of the 3'UTR had a strong affect on expression of a Ppo.L1925-encoded ORF. Changing the sequence of the last six nucleotides can prevent expression or increase it over ten-fold. These effects on expression levels were not seen in introns when the sequence of nucleotides 41-47 was changed (Lin and Vogt 2000).

The Ppo.L1925 5' end is important in splicing, as binding of this sequence and the 5' exon sequence with the intron's internal guide sequence (IGS) determines the location of the 5' splice site. The effect of the Ppo.L1925 5' end sequence on translation, though, has not been investigated. It is easy to imagine that the 5' end sequence could be important for translation, possibly by interacting with translation factors or by promoting the formation of a translation-competent RNA structure.

In this chapter, I describe experiments designed to examine the role of Ppo.L1925 sequences in splicing and expression, by introducing these sequences into the to chimeric Tth.L1925 intron. The first sequence I examine is the Ppo.L1925 3'UTR. In the previous chapter, I found that, unlike Ppo.L1925-based introns, chimeric Tth.L1925 did not require the 3'UTR sequence to splice properly. I now address how the 3'UTR sequence affects expression from chimeric Tth.L1925 introns. Does adding the 3'UTR reduce expression, and conversely do the mutant, high-expressing 3'UTR sequences increase expression, in Tth.L1925 as they appear to do in Ppo.L1925? The 3'UTR has been implicated as playing a role in the post-splicing processing of Ppo.L1925. Does the 3'UTR affect chimeric Tth.L1925 introns in the same way?

I also examine the Ppo.L1925 5' end sequence. How does this sequence affect translation from a chimeric Tth.L1925 intron? Does the 5' end sequence somehow interact with the 3'UTR sequence? I address these questions by introducing the Ppo.L1925 5' end sequence into chimeric Tth.L1925 introns.

Results

HEUTR sequences cause increased β-galactosidase activity in Ppo-α-HEUTR introns integrated in yeast

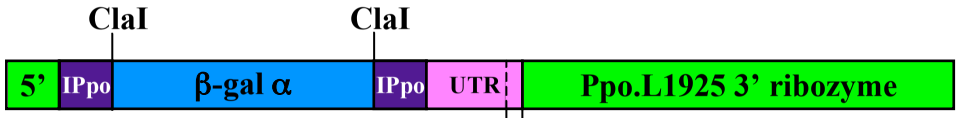
Nucleotides 48-53 of the 3'UTR play an important role in translation from Ppo.L1925-based intron RNAs through an unknown mechanism. Previously, Jue Lin created a pool of Ppo-α introns, Ppo-α ran 1, with random sequences at these six positions. The majority of yeast strains containing integrated introns from the Ppo-α ran1 pool exhibited much higher levels (over ten fold more) of β-galactosidase activity than those containing the wild type Ppo-α intron. Thus, it appears that changing nucleotides 48-53 of the Ppo.L1925 3'UTR can greatly increase expression, although no consensus for the high expressing 3'UTR (HEUTR) sequences from the

Ppo-α ran1 pool was apparent, and many of the introns had very different sequences in this region.

An alternate explanation for the high levels of β-galactosidase activity seen from the yeast with Ppo-α ran1 introns is that the increases are not caused by changes in the 3'UTR sequence, but by other mutations in the host yeast genome. To differentiate between these two possibilities, I selected six yeast strains containing individual, high expressing introns from the Ppo-α ran1 pool, which I named Ppo-α-HEUTR A through F (Figure 4.2). I amplified the intron and the surrounding rDNA sequence from these six isolates, cloned these fragments into the yeast shuttle vector pRS423 and trans-integrated them into yeast strain NOY505GalΩT. If the high β-galactosidase activity was due to a mutation in the original host strain, then it should not be carried over to the new host yeast and the freshly trans-integrated strains should not exhibit high levels of β-galactosidase activity. On the other hand, if the increase in expression level was due to the 3'UTR sequence, the freshly trans-integrated strains should exhibit the same high β-galactosidase activity as the original strains. I assayed the β-galactosidase activity of crude extracts of the intron containing yeast strains and found that all of the intron-integrated yeast exhibited the high β-galactosidase activities, with the highest, Ppo-αHEUTR C, exhibiting over 700 U, nearly 50 fold more than the wild type Ppo-α. This result confirms that the 3'UTR sequences are responsible for the increased β-galactosidase activity seen from yeast containing these introns.

Ppo.L1925 3'UTR sequences do not have a large effect on β-galactosidase activity from chimeric Tth-α introns

Since the 3'UTR sequence can have such a large effect on expression levels from Ppo-α, I decided to see what effect it would have on expression from Tth-α. In Ppo.L1925, the wild type UTR sequence is thought to have evolved to down regulate



Intron	Sequence (3'UTR nt 41-53)	β-galactosidase activity (U)
Ppo-α (WT 3'UTR)	**AAAACGCUGACCG**	15
Ppo-αHEUTR A	**AAUAGG**	430
Ppo-αHEUTR B	**GUUUAA**	340
Ppo-αHEUTR C	**CCUCGA**	770
Ppo-αHEUTR D	**UAAAUU**	500
Ppo-αHEUTR E	**GUUUAU**	240
Ppo-αHEUTR F	**GCUCAA**	230

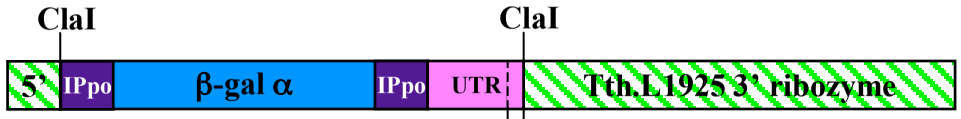
Figure 4.2: Ppo-αHEUTR mutants. These introns were cloned from their respective NOH505GalΩT strains and re-trans-integrated into NOY505GalΩT. The red letters in the second column indicate the nucleotide positions randomized in the intron pool. β-galactosidase activity in the third column is in Miller Units.

expression from the intron, as evidenced by the observation that majority of random UTR nucleotide 48-53 sequences resulted in an intron with increased expression. However, it is impossible to test what the base level of expression of an integrated intron would be without any 3'UTR, since deleting the this sequence prevents splicing.

Adding the 3'UTR sequence to the Tth-α intron could have several possible effects. First, and possibly least interestingly, the UTR sequence could disrupt splicing, which would prevent the intron from trans-integrating. Second, the 3'UTR sequence might decrease expression from the chimeric Tth.L1925 intron. Third, the 3'UTR sequence might promote expression from the chimeric Tth.L1925 intron.

To add the 3'UTR sequence to Tth-α, I used a two-step PCR process to fuse the 3'UTR sequence to the α ORF. I then cloned this PCR product into the ClaI site of Tth-ClaI to create the intron Tth-αUTR (Figure 4.3) and trans-integrated the Tth-αUTR intron into yeast strain NOY505GalΩT. I found that the majority of potential trans-integrants contained integrated introns (data not shown), indicating that the addition of the UTR sequence does not significantly disrupt Tth-αUTR splicing. The cells exhibited 5 U of β-galactosidase activity, somewhat lower than the 10 U produced from the Tth-α intron. This slight reduction in activity suggests that the 3'UTR sequence from *Physarum* does not have a major positive or negative effect on ORF expression in the *Tetrahymena* intron.

If the wild-type Ppo.L1925 3'UTR sequence does not have a large effect on Tth-α, how would the HEUTR sequences affect Tth-α? To address this question, I chose HEUTR sequences from Ppo-αHEUTR A, Ppo-αHEUTR B, and Ppo-αHEUTR C (Figure 4.3), to insert into Tth-α, thus creating introns Tth-αHEUTR A, Tth-αHEUTR B, and Tth-αHEUTR C. After trans-integrating the HEUTR introns, again I found that the majority of potential trans-integrants contained integrated introns (data



Intron	Sequence (3'UTR nt 39-53)	β-galactosidase activity (U)
Tth-αUTR	**GCAAAACGCUGACCG**	5
Tth-αHEUTR A	**AAUAGG**	4
Tth-αHEUTR B	**GUUUAA**	4
Tth-αHEUTR C	**CCUCGA**	4

Figure 4.3: Tth-αUTR and Tth-αHEUTR introns. Red letters in the second column indicate the nucleotide positions randomized in the intron pool. β-galactosidase activity in the third column is in Miller Units.

not shown), indicating that the HEUTR sequences does not disrupt splicing in the context of Tth-α. The three HEUTR introns exhibited similar levels of β-galactosidase activity, about 4 U, which is roughly the same as that of Tth-αUTR. Thus, the HEUTRs found previously to greatly increase ORF expression in the *Physarum* intron Ppo-α, do not function similarly in the context of Tth-α.

It is possible that the HEUTRs were being translated at a high level, but the UTR presence destabilized the intron RNA, which would reduce the message for and therefore, the protein product of, the Tth-αHEUTR introns. To address this possibility, I performed a northern blot (Figure 4.4). Using total RNA from yeast cells with integrated Tth-α, Tth-αUTR, the three Tth-αHEUTR introns, as well as Ppo-α and the three Ppo-HEUTR introns. The probe was an *in vitro*-made radiolabeled RNA complementary to Tth-α, which should hybridize with any α, Tth.L1925, and Ppo.L1925 ribozyme RNA. However, since the two entire ribozymes share only 70% sequence identity, the Ppo.L1925 intron was expected to give a weaker signal. No significant difference between the amount of full-length (i.e. potential message) Tth-α, Tth-αUTR, or Tth-αHEUTR RNA was evident. Thus, I conclude that neither the wild type 3'UTR nor HEUTR sequences significantly change the amount of α-encoding RNA being translated.

The northern blot also revealed at least two species of Tth-α RNA smaller than the full-length Tth-α. For comparison, in yeast, Ppo.L1925 RNA is known to be processed at two sites, internal processing sites one and two (IPS1 and IPS2). IPS1 processing occurs both *in vivo* and in *Physarum*, but IPS2 processing occurs only in yeast (Ruoff, Johansen et al. 1992; Lin and Vogt 1998). These processing sites are only 15 nucleotides apart and are both located in the P1 stem loop (Figure 4.7 A), with IPS1 cleavage occurring at the end of the 3'UTR and IPS2 cleavage occurring at the base of the P1 stem. In Ppo-α, all of the RNA species produced by cleavages are of

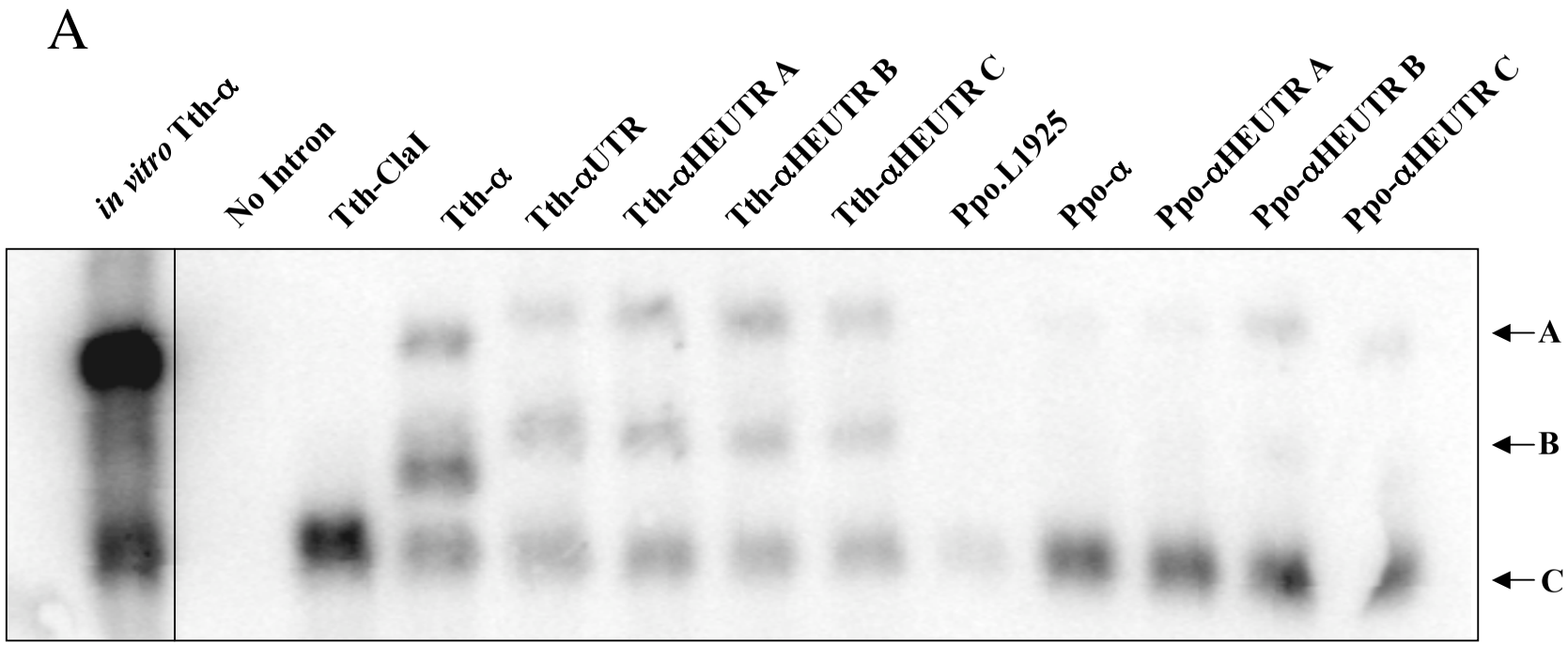


Figure 4.4: Northern blot of Tth-α and Ppo-α intron RNA. A. This northern blot was probed with an *in vitro* transcribed radiolabeled RNA that is complementary to the Tth-α sequence. Lane 1 0.1 µg of in vitro transcribed Tth-α RNA. Lane 2-13: 2.5 µg of total RNA isolated from a NOY505GalΩT strain with the listed intron integrated into its rDNA. Arrow A indicates full-length intron RNA, Arrow B indicates asymmetrically processed intron RNA, and Arrow C indicates RNA processed in the P1' stem.

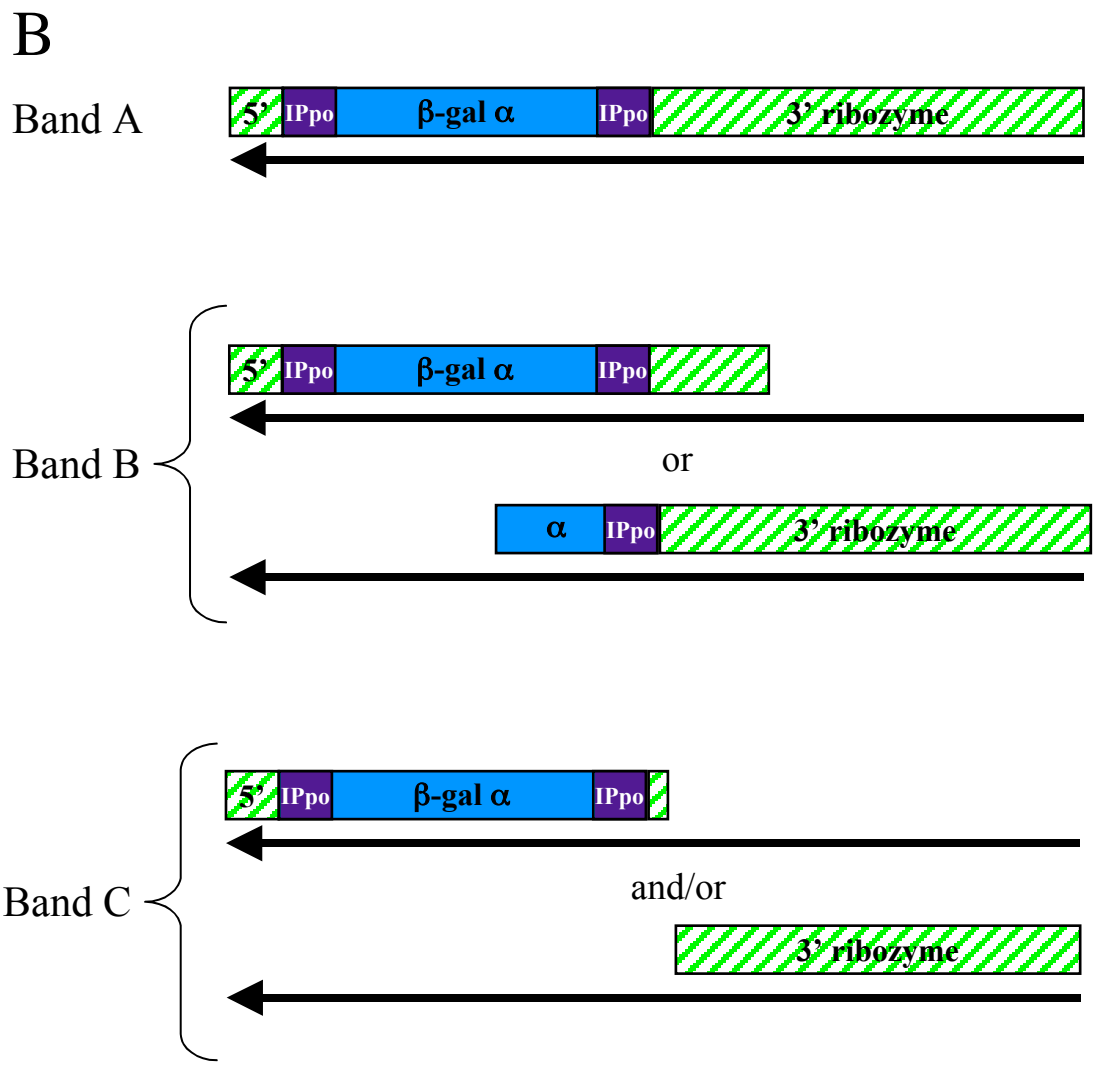


Figure 4.4 (Continued): The potential identities of the RNAs composing bands A, B, and C. The black arrow represents the anti-sense RNA probe. Band A consists of the full-length intron RNA. Band B consists of asymmetrically-cleaved RNA. Band C consists of a symmetrically-cleaved RNA.

similar size, and thus run at the same position on a northern blot (Figure 4.4, Band C). All of the Tth-α-based introns exhibit a band of similar size, which is indicative of processing at an IPS1/IPS2-like site.

Because the products of cleavage at IPS1 and IPS 2 are so similar in size, it is impossible to tell from this northern blot if only one cleavage is occurring in Tth-α introns or if, like their cousins from *Physarum*, they are undergoing two cleavages. I performed a primer extension reaction to more accurately map the central Tth-α processing site, using the same radio-labeled primer in a parallel sequencing reaction. Both reactions were then run on the same denaturing poly-acrylamide sequencing gel. Primer extension of total RNA extracts from yeast cells with integrated Tth-α, Tth-αUTR, and Tth-αHEUTR A produced two major bands (Figure 4.5). These bands correspond to a G residue in the second ClaI sequence, just 3' of the α ORF, and a G at the base of the P1 stem. Since group I intron processing occurs by a G addition mechanism (see Figure 1.2) (Ruoff, Johansen et al. 1992; Lin and Vogt 1998), the primer extension bands should be one nucleotide longer than the sequencing band that corresponds to the actual cleavage site. Thus, it appears that Tth-α introns have two processing sites in the P1 stem region, IPS1 between Tth-α nucleotides 314-315 and IPS2 between Tth-α nucleotides 329-330. These two processing sites correspond closely with IPS1 and IPS2 of Ppo.L1925.

Tth-α also undergoes cleavage at a third site, IPS3 (Figure 4.4, Band B). The resulting RNA species runs between the full-length intron and IPS1/IPS2 processed RNAs. Since it is larger than the IPS1/IPS2 RNA, it must be the result of an asymmetric cleavage of the RNA, placing IPS3 either within the α ORF RNA or within the ribozyme RNA. A band corresponding to the complementary fragment is not visible on the blot, and could have run off of the gel. Alternatively, the smaller fragment may not be stable.

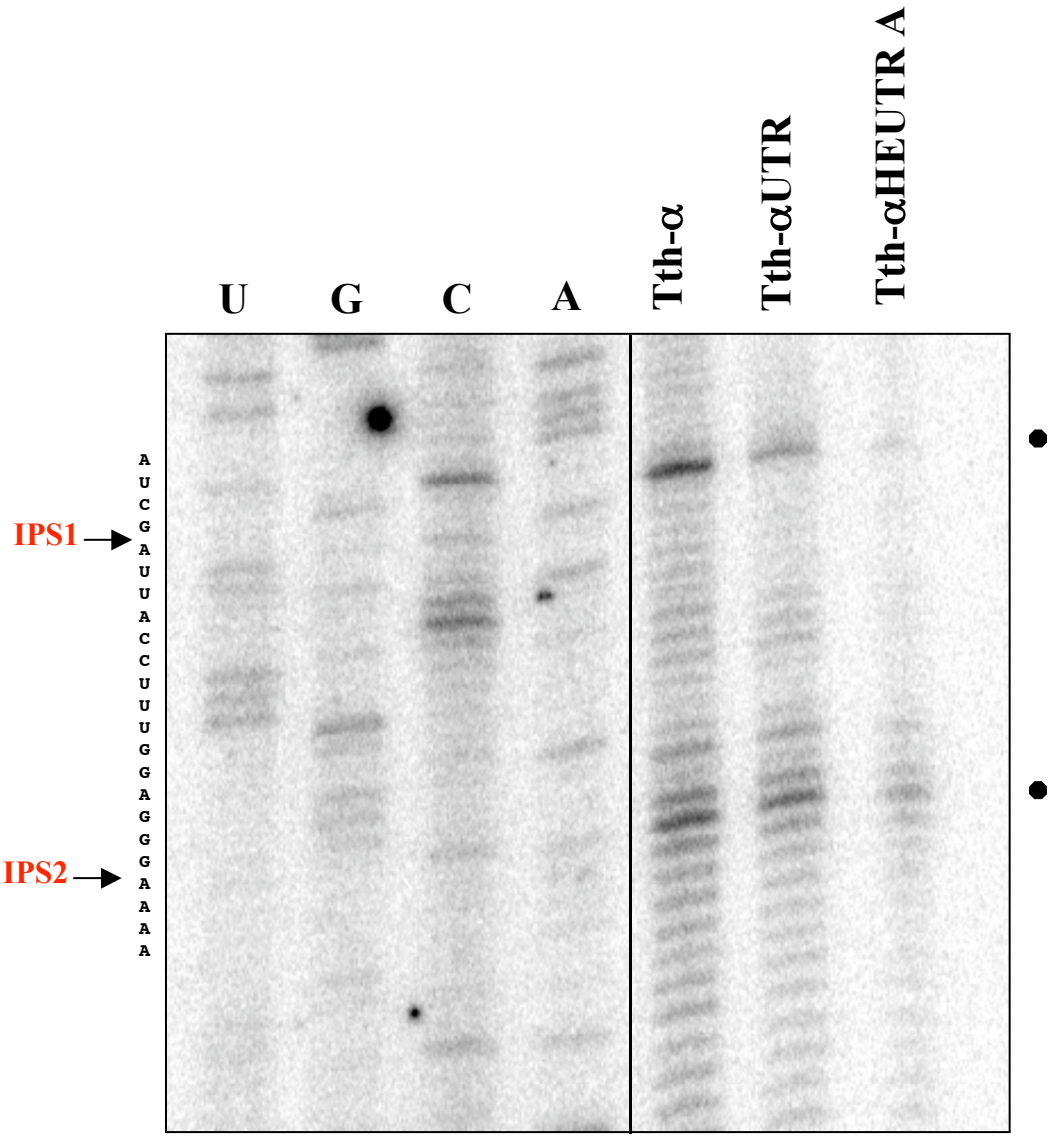


Figure 4.5. Primer extension analysis of the Tth-α P1' region. Sequencing and primer extension analysis of Tth-α, Tth-αUTR, and Tth-αHEUTR A.

In order to identify the IPS3-derived band, I prepared a second northern blot using two different probes sequentially. The first was an end-labeled oligonucleotide complementary to the 3' end of the Tth.L1925 intron. The second was an end-labeled oligonucleotide complementary to the 5' end of the α sequence. After stripping off the first probe no bands remained visible (data not shown). As seen in Figure 4.6, the 3' end probe bound to the full-length and both processed RNAs, while the 5' end probe bound only to the full-length intron RNA. This result implies that the processing site is within the α sequence and not the ribozyme. Interestingly, the only species of intron RNA that contains the α ORF visualized in either of the blots is the full-length intron. It is possible that the ribozyme structure protects RNA species that contain it, while processed pieces lacking the 3' end are degraded.

In summary, based on these experiments, the Tth-α introns process at three sites (Figure 4.7, two that are similar to Ppo.L1925 processing sites IPS1 and IPS2, and a third site, IPS3, that is in the α ORF. None of the processed RNA fragments that contain the α ORF appear to be stable, and so I conclude that it is unlikely that processed RNA species do not contribute to the amount of α being produced in the cell. Based on these results, however, it is impossible to rule out the possibility that undetected processed fragments, or another, low-quantity intron-derived RNA species is acting as the message for intron-encoded ORF translation.

The 5' end of Ppo.L1925 increases β-galactosidase activity from Tth-α

Unlike the 3'UTR, the effect of the intron 5' end sequence on translation has not been examined. This sequence, between the 5' end of the intron and the I-PpoI start codon, is only twelve nucleotides long. Eleven of these nucleotides are encoded by the intron and one is the exogenous G residue added during the first step of splicing. Even in this small sequence, Ppo-α and Tth-α show large differences. The Ppo-α 5' end is C-rich, (6 out of 13 residues, while the Tth.L1925 sequence is A rich

4.6 Northern blot of Tth-α and Ppo-α intron RNA with 5' and 3' end specific probes. The top northern blot was probed with an end-labeled oligonucleotide primer that is complementary to the 3' region of the Tth.L1925 intron. The bottom northern blot was probed with an end-labeled oligonucleotide complementary to the 5' end of the β-galactosidase α fragment. Lane 1 0.1 μg of in vitro transcribed Tth-α RNA. Lane 2-13: 2.5 mg of Total RNA isolated from a NOY505GalΩT strain with the listed intron integrated into the rDNA. Arrow A indicates full-length intron RNA, Arrow B indicates asymmetrically processed intron RNA, and Arrow C indicates RNA processed in the P1' stem.

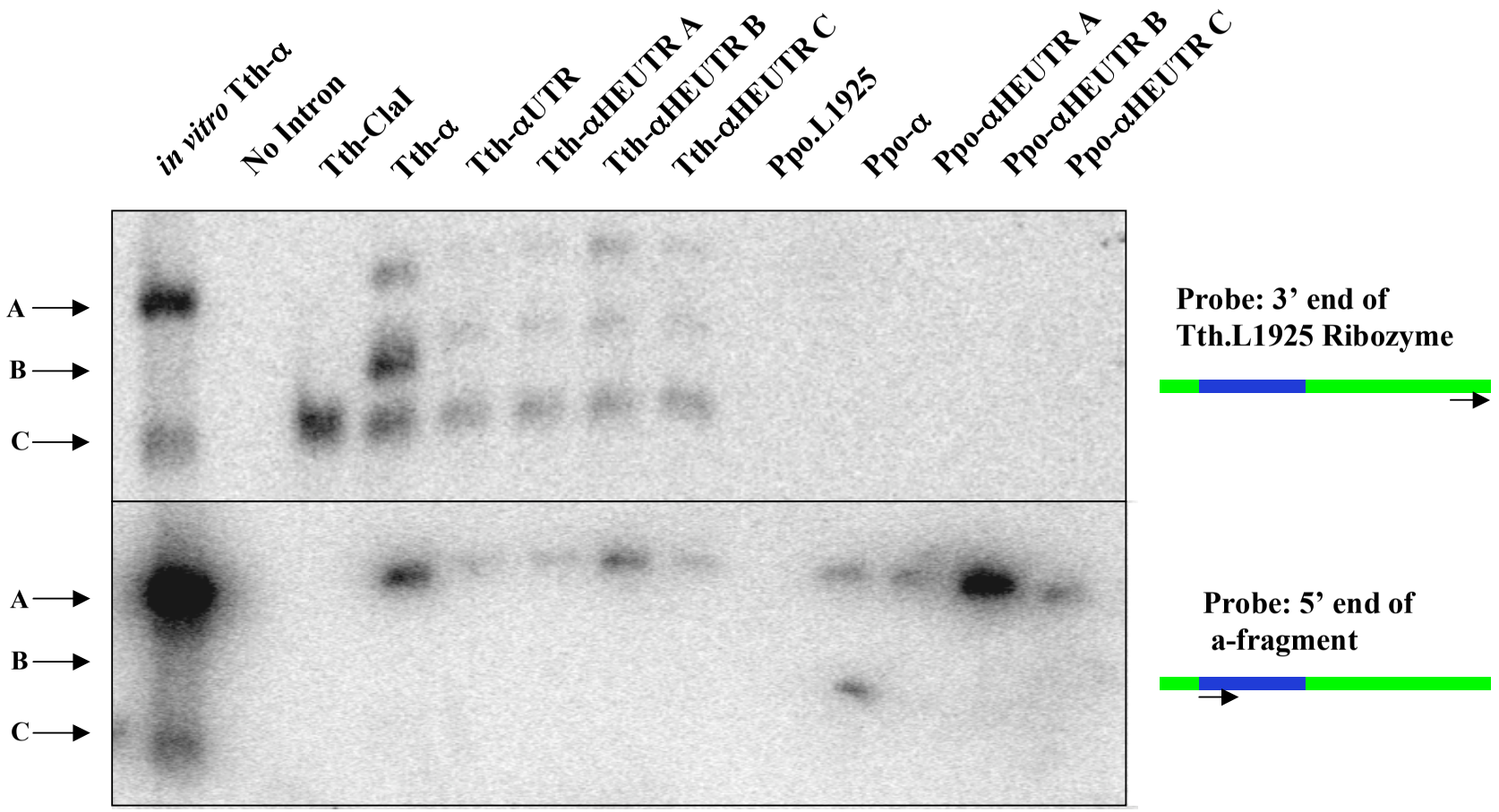
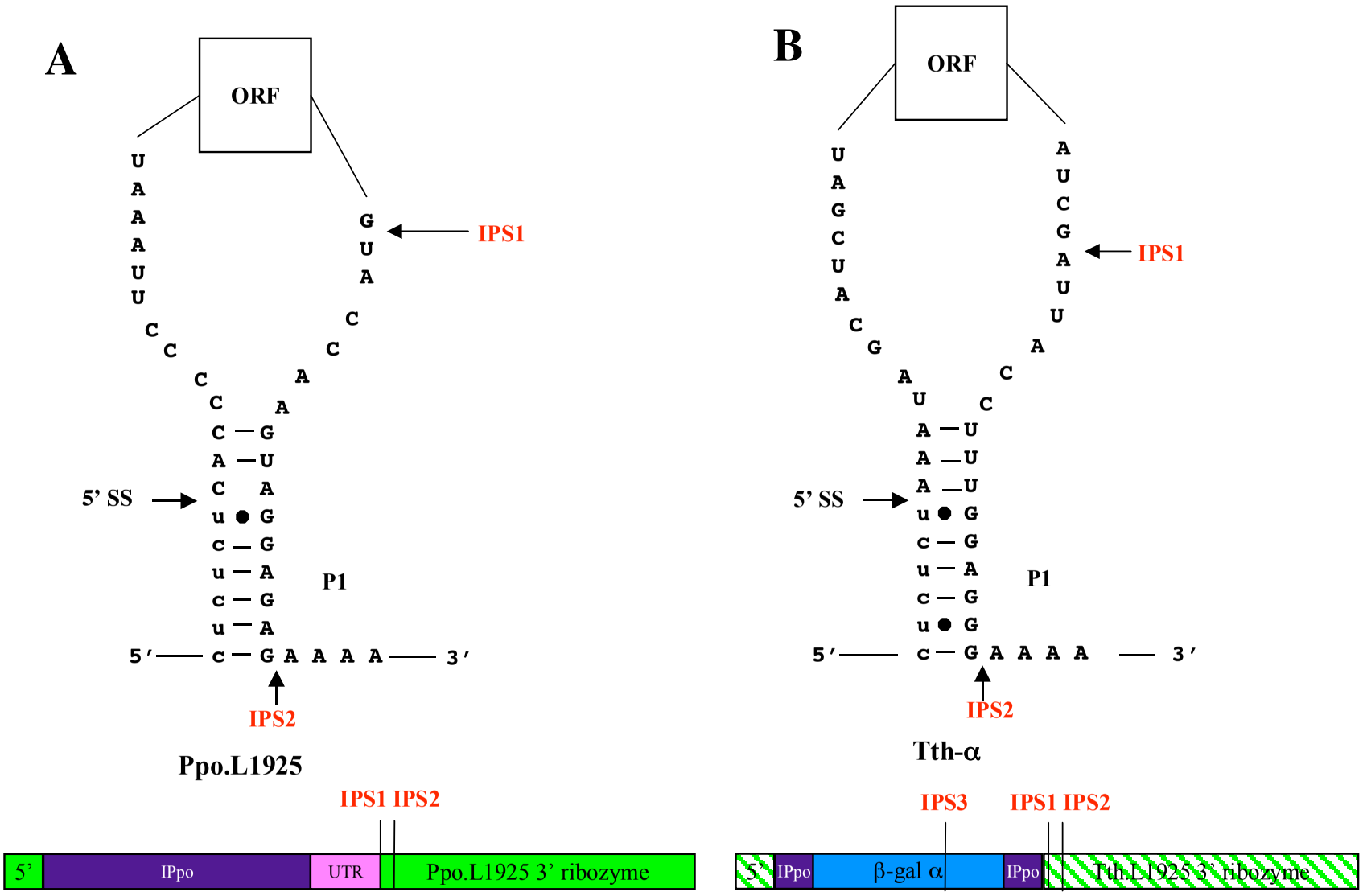

in vitro Tth-α
No Intron
Tth-ClaI
Tth-α
Tth-αUTR
Tth-αHEUTR A
Tth-αHEUTR B
Tth-αHEUTR C
Ppo.L1925
Ppo-α
Ppo-αHEUTR A
Ppo-αHEUTR B
Ppo-αHEUTR C
A
B
C
A
B
C
Probe: 3' end of
Tth.L1925 Ribozyme
Probe: 5' end of
a-fragment

Figure 4.7: A comparison of the internal processing sites of Ppo.L1925 and Tth-α. A. Ppo.L1925 is processed at two points, IPS1 at the end of the 3'UTR and IPS2 at the base of the P1' stem. B. Tth-α is processed at three points. IPS1 occurs in the 3' ClaI site, IPS2 occurs at the base of the P1' stem and IPS3 occurs with the alpha ORF.


A
ORF
IPS1
5' SS
P1
IPS2
Ppo.L1925
IPS1 IPS2
5'
IPpo
UTR
Ppo.L1925 3' ribozyme
B
ORF
IPS1
5' SS
P1
IPS2
Tth-α
IPS3
IPS1 IPS2
IPpo
β-gal α
IPpo

(6 out of 13 residues) (Figure 4.8 A). The first three nucleotides of both introns are very important for splicing, as these nucleotides base-pair with the IGS, which determines the 5' splice site (Figure 4.8 B) (Doudna, Cormack et al. 1989). It is difficult to predict how exchanging the Ppo.L1925 5' end for the Tth-α 5' end would affect translation. Since Tth.L1925 does not contain an ORF sequence in *Tetrahymena*, it presumably has not been under any selective pressure for a 5' end sequence that would affect translation. The potential function of the 5' end in Ppo.L1925 is also unknown. It might function to promote translation of the downstream ORF or to down-regulate the translation of the potentially toxic I-PpoI protein, or it might have no function in translation

In order to simplify the addition of the Ppo.L1925 5' end to chimeric Tth.L1925 introns, I modified the Tth-ClaI intron to make Tth-5'P-ClaI (Figure 4.9). Tth-5'P-ClaI has the Tth.L1925 5' end sequence replaced with that of Ppo.L1925. I also changed the complementary nucleotides in the P1 stem to preserve the IGS contacts in order to minimize potential disruption of the splicing reaction. These changes, here referred to as 5'P (for the **5'** end of ***P****hysarum*), appear not to have a negative effect on intron splicing, as Tth-5'P-ClaI is able to trans-integrated into yeast (data not shown).

The Tth-5'P-ClaI intron allowed me to effectively add the 5'P sequence to Tth-α by cloning the α ORF from Tth-α into the ClaI site of Tth-5'P-ClaI site to create intron Tth-5'P-α. Like its parent introns, Tth-5'P-α was able to trans-ingrate into yeast strain NOY505GalΩT, indicating that the Ppo.L1925 5' end sequence is compatible with intron splicing. β-galactosidase assays of yeast crude extracts showed 80 U of activity, almost a tenfold increase over the 10 U of activity exhibited by Tth-α (Figure 4.10 A).

Figure 4.8: Comparison of the 5' end sequences of Ppo.L1925 and Tth-α. A. An alignment of the Ppo.L1925 and Tth-alpha 5' ends. Green bases indicate positions where there are differences in the nucleotide sequence. The orange bases indicate the ORF start codon. B. The 5' end secondary structure.

A

Ppo.L1925 5' end: CACCCCCUUAAAUATG...
Tth-α 5' end: AAAUAGCAUCGAUATG...

B

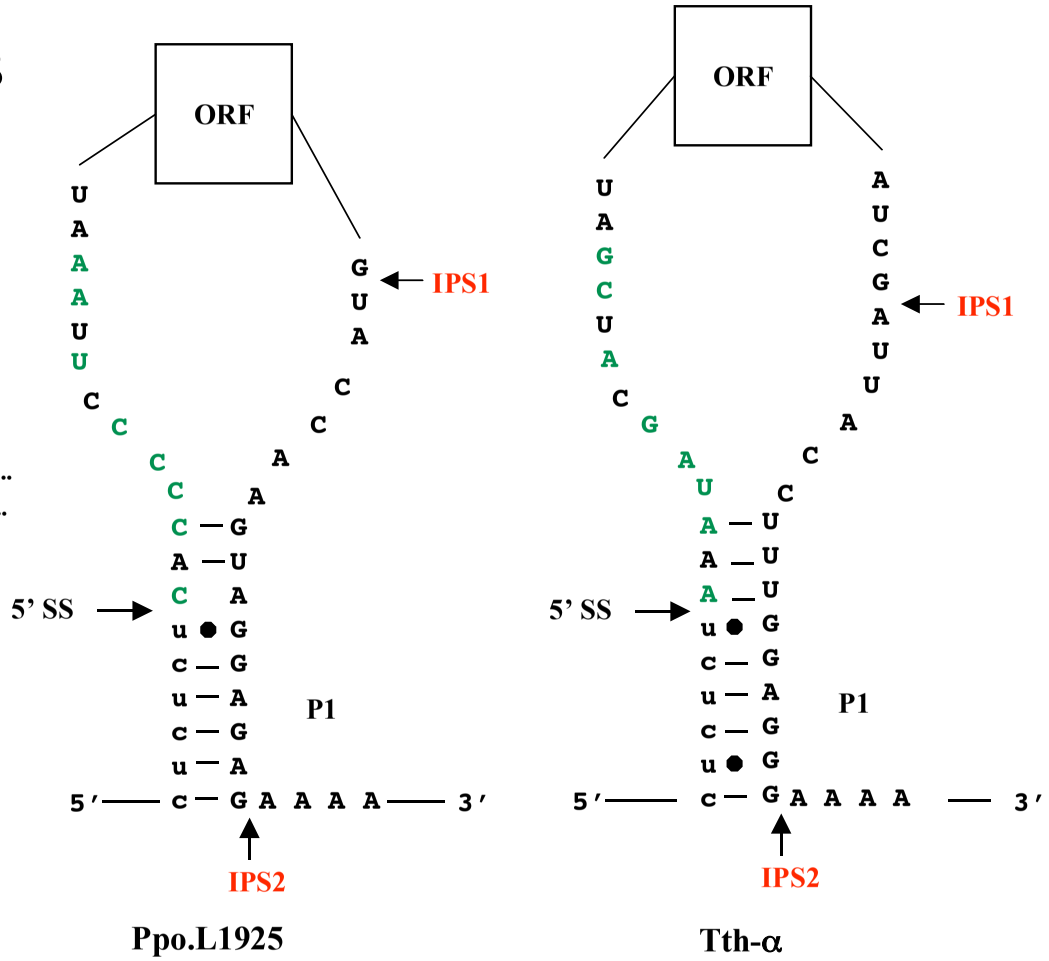
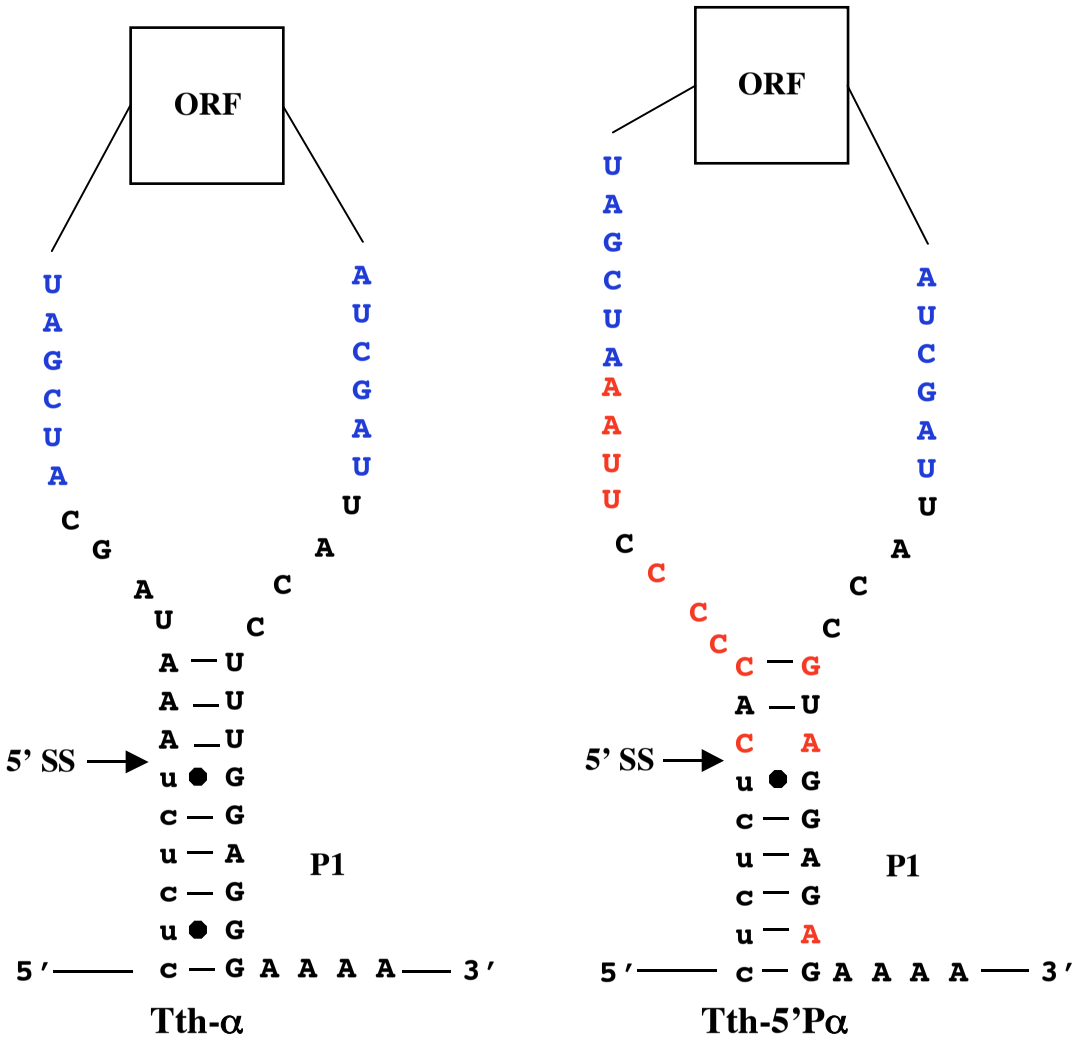
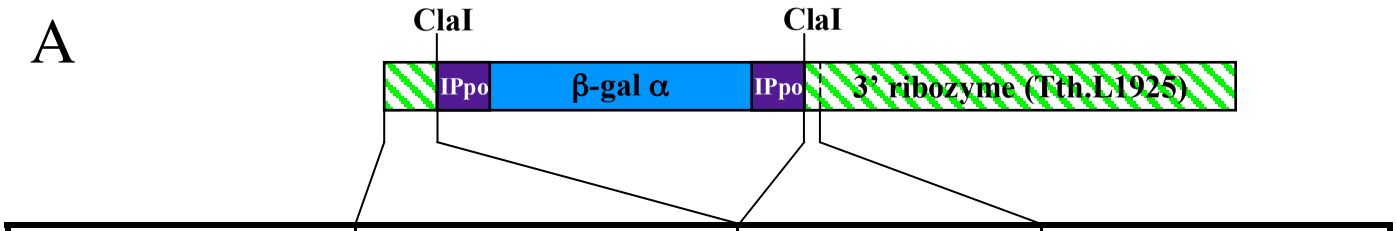


Figure 4.9: The P1 regions of Tth-α and Tth-5'P-α. Tth-5'P-α nucleotides changed from Tth-alpha are shown in red and nucleotides composing the ClaI sites are shown in blue. Intron sequence is in caps while the 5' exon sequence is in lower case.


ORF
U A G C U A
A U C G A U U
C G A C A U C
A — U
A — U
A — U
5' SS
u ● G
c — G
u — A
c — G
u ● G
P1
5' c — G A A A A 3'
Tth-α
ORF
U A G C U A A A U U C C C C C
A U C G A U U
C G A C A U C
C — G
A — U
C A
5' SS
u ● G
c — G
u — A
c — G
u — A
P1
5' c — G A A A A 3'
Tth-5'Pα

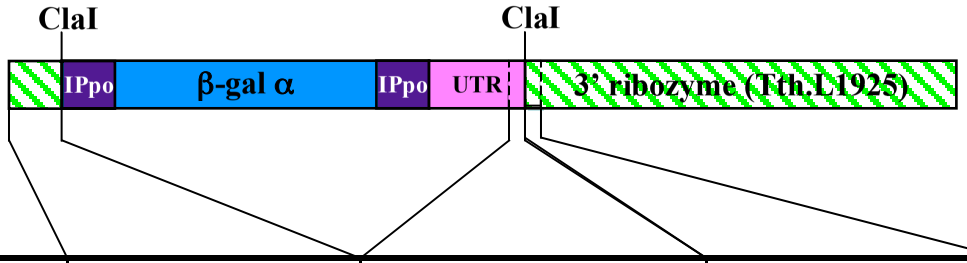
A



Intron	Sequence (5' end)	Sequence (P1')	β-galactosidase activity (U)
Tth-α	AAAUAGC	UACCUUUGGAGGG	10
Tth-5'P-α	CACCCCCUUAA	UACCGUAGGAGAG	80

Figure 4.10: Tth-5'P-α introns. A. Tth-α and Tth-5'P-α. Red letters in the second and third columns indicate the nucleotides that are not shared between Tth-α and Tth-5'P- α. All β-galactosidase activity is in Miller Units.

B



Intron	Sequence (5' end)	Sequence (3'UTR nt 39-53)	Sequence (P1')	β-galactosidase activity (U)
Tth-5'P-αUTR	CACCCCCUUAA	GCAAAACGCUGACCG	UACCUUUGGAGGG	3
Tth-5'P-αHEUTR A	CACCCCCUUAA	AAUAGG	UACCUUUGGAGGG	10
Tth-5'P-αHEUTR B	CACCCCCUUAA	GUUUAA	UACCUUUGGAGGG	30
Tth-5'P-αHEUTR C	CACCCCCUUAA	CCUCGA	UACCUUUGGAGGG	5

Figure 4.10: (Continued) B. Tth-5'P- α-UTR and Tth-5'P- α HEUTR introns. Red letters in the second, third, and fourth columns indicate the nucleotides that are not shared between introns. All β-galactosidase activity is in Miller Units.

The 5'P sequence might interact with the 3'UTR . To test this hypothesis, I cloned the α-3'UTR sequence from Tth-αUTR into the ClaI site of Tth-5'P-ClaI to make intron Tth-5'P-αUTR (Figure 4.10 B). Yeast cells with integrated Tth-5'P-αUTR exhibited only 3 U of β-galactosidase activity. Thus, the 3'UTR sequence strongly down regulates expression from Tth-5'P-αUTR.

As documented earlier in this chapter, the *Physarum* HEUTR sequences recovered from Jue Lin's randomization experiment do not increase expression from Tth-α as they do in Ppo-α. It is possible that those sequences could function in an intron with the 5'P sequence. To create the introns to test this, I cloned the α-UTR fragment from the appropriate Tth-αHEUTR intron into the ClaI site of Tth-5'P-ClaI to make introns Tth-5'P-αHEUTR A, B, and C, respectively (Figure 4.10). Yeast cells with these introns integrated exhibited higher of β-galactosidase activity than those with the corresponding 5'P-lessTth-α-HEUTR intron. The HEUTR intron with the highest activity, Tth-5'P-alpha-HEUTR B, exhibited 30 U of activity, nearly tenfold more than the 5'P-lacking Tth-α-HEUTR B intron. However, this is still under half of the activity of the UTR-lacking Tth-5'P-α and these introns did not exhibit the very high levels of activity seen from the Ppo-αHEUTR introns, suggesting that the HEUTR sequences do not functionally interact with the 5' end in Ppo-α. Instead, it seems that all the 3'UTR sequences lower expression, both from Tth-α and from Tth-5'P-α introns.

Does the 5'P sequence increase expression of other chimeric introns besides Tth-α? The 5' end sequence could be interacting with the α ORF to increase expression, even though there are no obvious regions where base-pairing could occur. On the other hand, perhaps the 5' ends sequence is interacting with downstream ribozyme sequences that a larger inserted ORF would discourage it from finding. In order to address this question, I created intron Tth-5'P-His3HA by cloning the

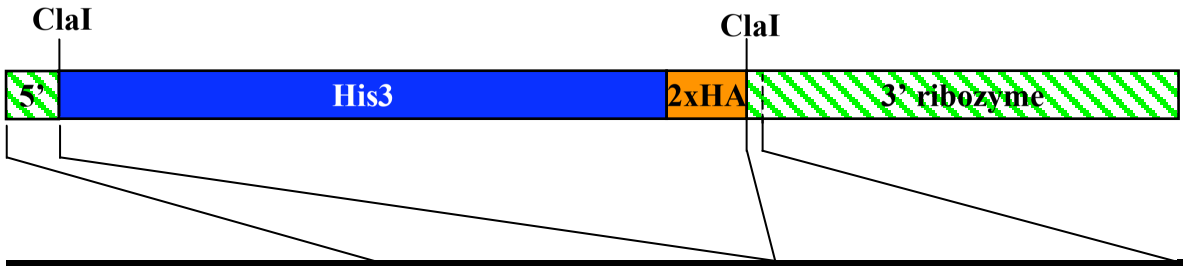
His3HA fragment of Tth-His3HA into the ClaI site of Tth-5'P-ClaI (Figure 4.11 A). I prepared total protein extracts from yeast with integrated Tth-His3HA and Tth-5'P-His3HA introns and compared them on a western blot (Figure 4.11 B). The 5' end sequence did not cause Tth-5'P-His3HA to produce more His3HA protein; instead the levels of His3HA were similar. Because of this result, I chose to continue work with Tth-alpah5'P to attempt to determine how the 5' end increases expression from that intron.

Use of a randomization approach to study the intron 5' end in expression of chimeric-Tth.L1925 introns

I decided to utilize a randomization approach, similar to that used previously to study the 3' UTR (Lin and Vogt 2000), to obtain a greater understanding of how the 5' end sequences affects splicing. In this approach a pool of Tth-5'P-α introns is created with a random nucleotide sequence at the 5' end. This pool is then trans-integrated into yeast and individual colonies are screened for β-galactosidase activity through growth on plates containing X-gal. This initial screen allows an estimation of the percentage of 5' end sequences that cause high, low, or no expression. If the majority of 5' end sequences cause Ppo-5'P-α-like expression levels, the majority of yeast screened should exhibit a dark blue phenotype. If increased expression requires sequences similar to the 5'P sequence, the majority of yeast should be light blue or white, and a dark blue phenotype should be rare. After the initial screen, individual dark blue, light blue, and white yeast colonies can be picked, their introns sequenced, and the β-galactosidase expression levels determined quantitatively. The sequences can then be analyzed for any motifs that might be associated with high or low expression levels.

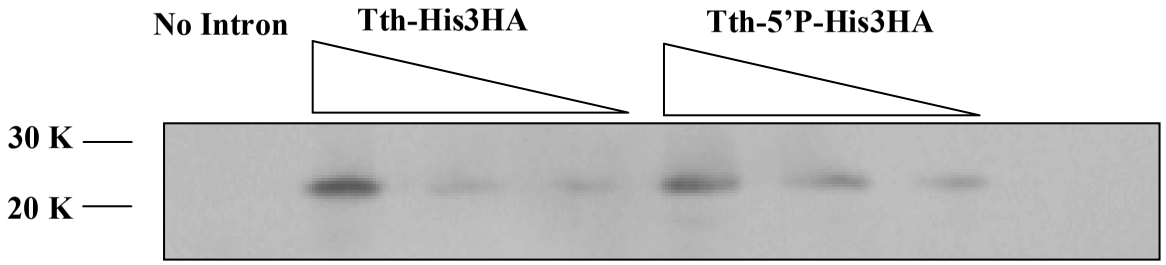
Figure 4.11: Tth-5'P-His3HA. A. Schematic representation of the Tth-His3HA and Tth-5'P-His3HA introns. Red letters in the second and third columns indicate the nucleotides that are not shared between introns. B. Western blot analysis of whole protein extracts from intron-containing NOY505GalΩT. Lane 1, no intron, 25 µg total protein, Lanes 2-4, Tth-His3HA 25 µg, 6.25 µg, and 2.5 µg total protein, respectively, and Lanes 5-7, Tth-5'P-His3HA 25 µg, 6.25 µg, and 2.5 µg total protein, respectively

A



Intron	Sequence (5' end)	Sequence (P1')
Tth-His3HA	AAAUAGC	UACCUUUGGAGGG
Tth-5'P-His3HA	CACCCCCUUAA	UACCGUAGGAGAG

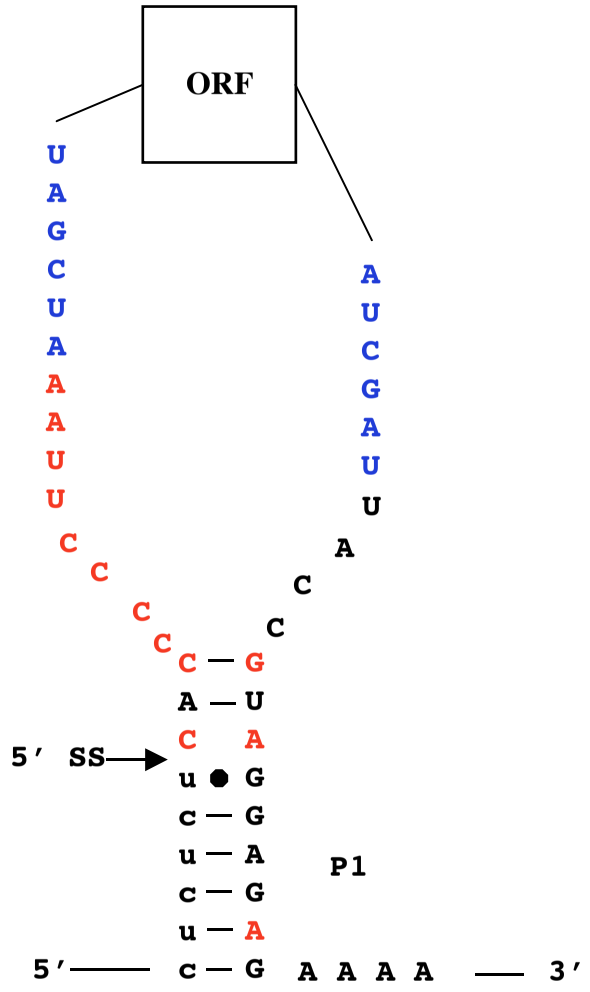
B

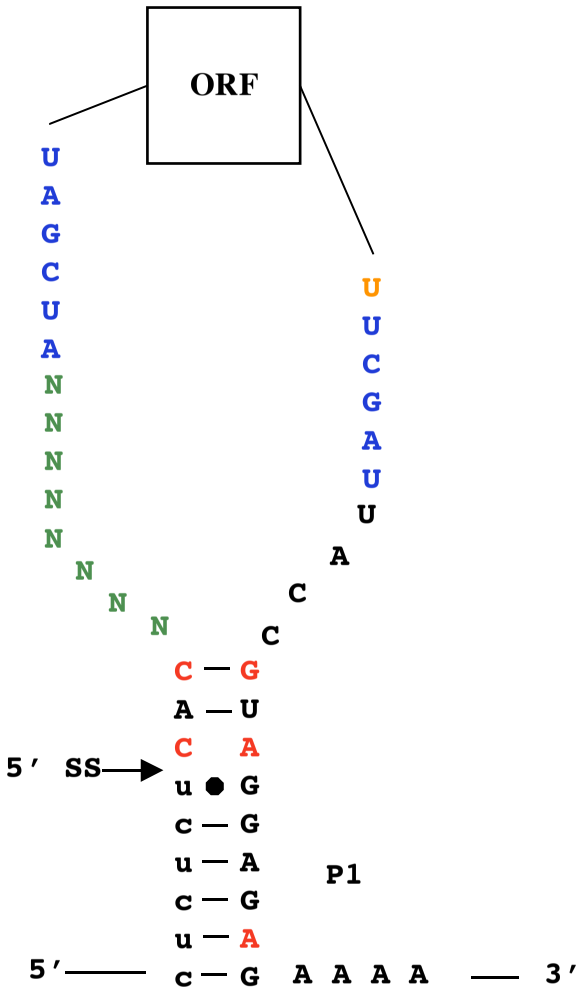


I decided to examine a pool of introns, called Tth-5'P-α-ran, with eight randomized nucleotides at Ppo-5'P-α positions 4-11 (Figure 4.12). The sequence complexity of this pool is 65,536, and hence only a small fraction of the different sequences could be analyzed. The first three nucleotides of the intron were left constant as they are involved in the P1 stem and changing them could negatively impact splicing.

The process used to create this pool is outlined in Figure 4.13. The intron pool was made by combining a pool of intron 5' end fragments with randomized fragments created through PCR with a plasmid encoding a version of the Tth-5'P-a intron which had its 5' end replaced with the kanamycin resistance gene, Kan. The randomized PCR fragment was then cloned into the plasmid to replace the Kan stuffer fragment. To estimate the number of unique introns encoded by the plasmid pool, I plated small aliquots of the E. coli transformation on 2YT+Amp plates and counted the resulting colonies. From these plates, the size of the plasmid pool was estimated to be roughly 50,000-100,000 unique transformants, which means that the pool should contain most of the possible 65,536 sequences. Additionally, the Kan stuffer fragment allows the estimation of the amount of uncut or relegated vector in the plasmid pool, as E. coli containing these plasmids are resistant to Kan, while E. coli containing one of the Tth- 5'P-α-ran plasmids are not. By plating small aliquots of the E. coli transformation on 2YT +Amp +Kan, I estimate that 9% of the plasmids still contained the Kan stuffer fragment. This is an acceptable number, as it means the majority of plasmids in the pool contain Tth-5'P-α-ran introns. As a last check of the integrity of the plasmid pool, I had a small aliquot of the pool sequenced. The electropherogram indicated roughly equal concentrations of A, C, G, and T at positions 4-11, implying that there was no large sequence bias in the plasmid pool.

Figure 4.12: Randomization of the Tth-5'P-α 5' end. Nucleotides in green are randomized in the Tth-5'Palpha-ran pool. Nucleotides involved in forming the P1 stem were unchanged to prevent disruption of intron splicing. The mutation that disrupts the 3' ClaI site is shown in orange. Nucleotides composing the ClaI sites are shown in blue. Intron sequence is in caps while the 5' exon sequence is in lower case.


ORF
5′ SS
P1
5′
3′
Tth-5’Pα


ORF
5′ SS
P1
5′
3′
Tth-5’Pα-ran

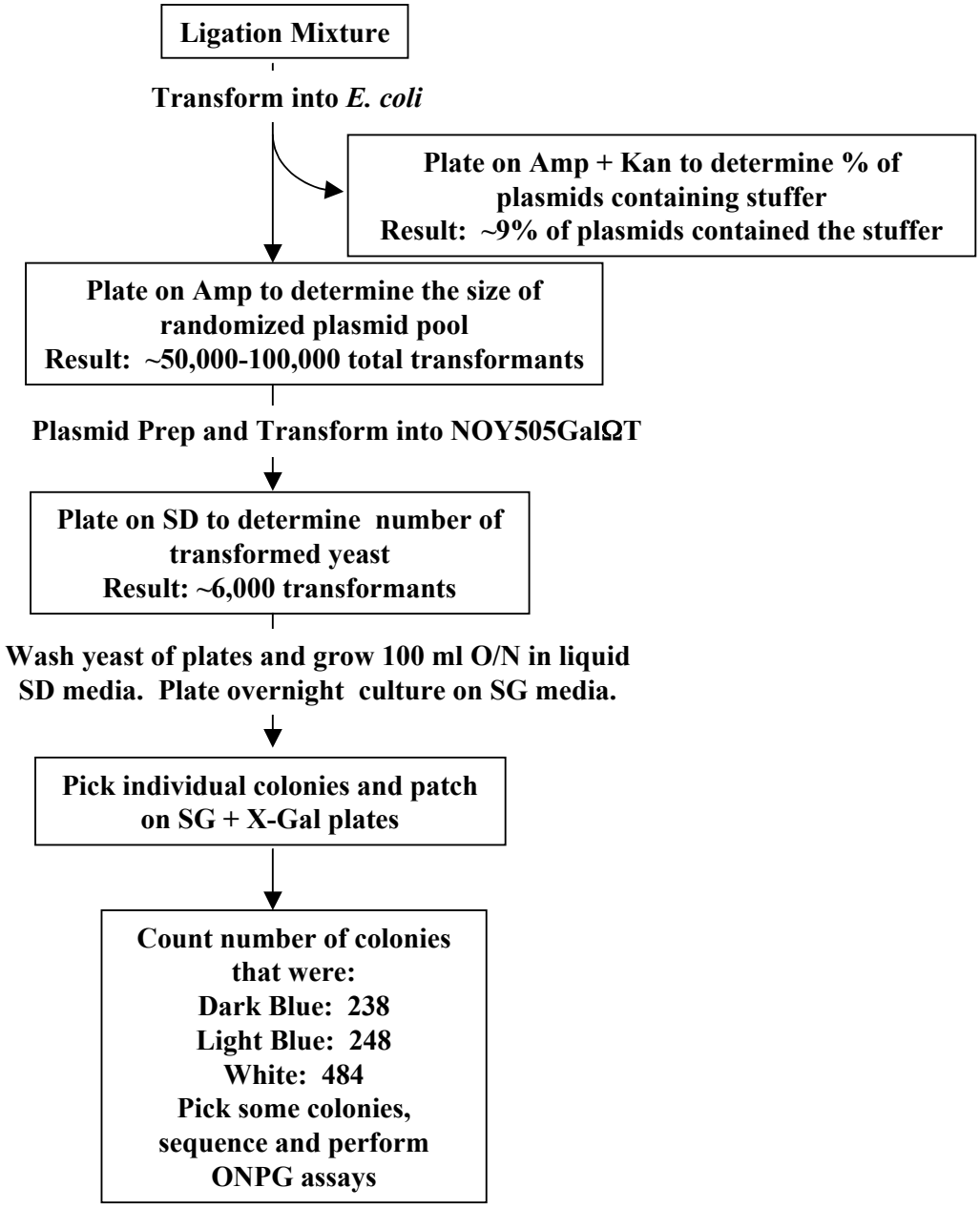


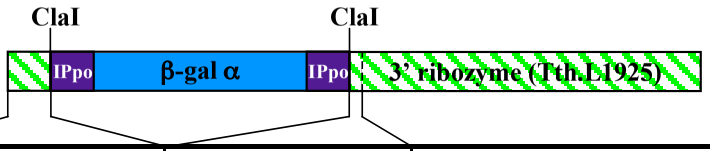
Figure 4.13: Process for creating a pool of yeast cells with integrated Tth-5'P-α-ran introns.

Confident that the plasmid pool contained a good representation of the possible Tth-5'P-α-ran sequences, I trans-integrated the Tth-5'P-α-ran pool into yeast, for which it was necessary to modify some of the procedures. The first step of trans-integration, co-transformation of the Tth-5'P-α-ran plasmid pool and pCpIPpo, was unchanged. I transformed yeast strain NOY505GalΩT and plated potential transformants on SD –His -Ura media. Counting the colonies on these plates revealed that there were ~6,000 yeast transformants. If I were trans-integrating a single intron, at this point I would have picked a single colony and streaked it on SG –His –Ura media to induce I-PpoI production. Since I wanted to trans-integrate all of ~6,000 introns in the individual colonies, I washed the yeast off of the SD plates, pooled them, incubated them in liquid SD media over night, and then plated them on SG –His –Ura media. Each yeast colony growing on the resulting plates should contain one of these ~6,000 introns or an I-PpoI resistance mutation.

To screen the level of β-galactosidase activity of introns from the pool, I picked ~1000 individual colonies, patched them on SG+X-gal plates and then scored the patches as either white, light blue, or dark blue. Of the scored colonies, 238 were dark blue, 248 were light blue, and 482 were white, implying that a large number of 5' end sequences achieve Tth-5'P-α levels of expression.

About ~25% of the 5' end sequences exhibited a high expression phenotype. In order to link individual sequences to levels of expression, I picked 21 individual colonies, PCR-amplified and sequenced their introns, and, in addition, conducted β-galactosidase activity assays on them. The results of the sequencing and β-galactosidase assays are shown in Figure 4.14. Many of the sequences are AT rich, although no obvious common motif is associated with either high or low expression. β-galactosidase activity ranges from 250 U, more than double that of Tth-5'P-α, to

Figure 4.14: Tth-5'P-α-ran intron 5' end sequence and β-galactosidase activity. Red letters in the second column indicate the wild type Tth-5'P-α sequence and green letters in the second column indicate nucleotides in the randomized region. Red letters in the third column indicate the nucleotides that are not shared between Tth-α and Tth-5'P-α.



Intron	Sequence (5' end)	Sequence (P1')	β-galactosidase activity (U)
Tth-5'Pα	CACCCCCUUAA	UACCGUAGGAGAG	80
Tth-5'α-ran 1	CACACUAUAUA	UACCGUAGGAGAG	250
Tth-5'α-ran 2	CACAUUAAAC	UACCGUAGGAGAG	240
Tth-5'α-ran 3	CACAAAACAAA	UACCGUAGGAGAG	200
Tth-5'α-ran 4	CACUAUAUACA	UACCGUAGGAGAG	190
Tth-5'α-ran 5	CACGAAAUAAC	UACCGUAGGAGAG	170
Tth-5'α-ran 6	CACAAGCAUAU	UACCGUAGGAGAG	170
Tth-5'α-ran 7	CACAGAACUAN	UACCGUAGGAGAG	170
Tth-5'α-ran 8	CACAUUUAGAU	UACCGUAGGAGAG	150
Tth-5'α-ran 9	CACAUCUAAAU	UACCGUAGGAGAG	130
Tth-5'α-ran 10	CACAAGAUUCC	UACCGUAGGAGAG	120
Tth-5'α-ran 11	CACGAUACUAG	UACCGUAGGAGAG	100
Tth-5'α-ran 12	CACACAAGGA	UACCGUAGGAGAG	80
Tth-5'α-ran 13	CACGUCGAAAU	UACCGUAGGAGAG	80
Tth-5'α-ran 14	CACCAUCGAU	UACCGUAGGAGAG	40
Tth-5'α-ran 15	CACGACGUCCC	UACCGUAGGAGAG	25
Tth-5'α-ran 16	CACUUAAUGUA	UACCGUAGGAGAG	20
Tth-5'α-ran 17	CACAAUGAAUA	UACCGUAGGAGAG	3
Tth-5'α-ran 18	CACGUGGCACC	UACCGUAGGAGAG	1

1 U, which is only slightly above the background activity seen yeast with no integrated introns. A few (Tth-5'P-α-ran 2, 12, and 14) have only seven nucleotides instead of eight in the randomized region. This may be due to errors in the synthesis of the 3' oligonucleotide used as a primer to make the introns. Having one less nucleotide in the randomized region has no noticeable effect on expression.

To determine if there is a specific folding motif, I analyzed high and low expressing sequences using two RNA secondary structure prediction programs, Kinefold (Xayaphoummine, Bucher et al. 2005) and RNAfold from the Vienna software package (Gruber, Lorenz et al. 2008). Since the Kinefold program has a 400 nucleotide limit, it could only be used to examine the P1 stem and the α ORF, while the RNAfold program, which has a 10,000 nucleotide limit, could be used to examine both the P1 stem and the α ORF and the entire intron sequence. I examined five high expressing intron (Tth-5'P-α and Tth-5'P-α ran 1-4) and five low expressing intron (Tth-α and Tth-5'P-α ran 15-18) sequences. The analysis focused on three structural factors that could be influencing levels of translation: if they shared folding motifs, if the 5' end of the intron was involved in a structure, and if the α start codon was involved in a structure.

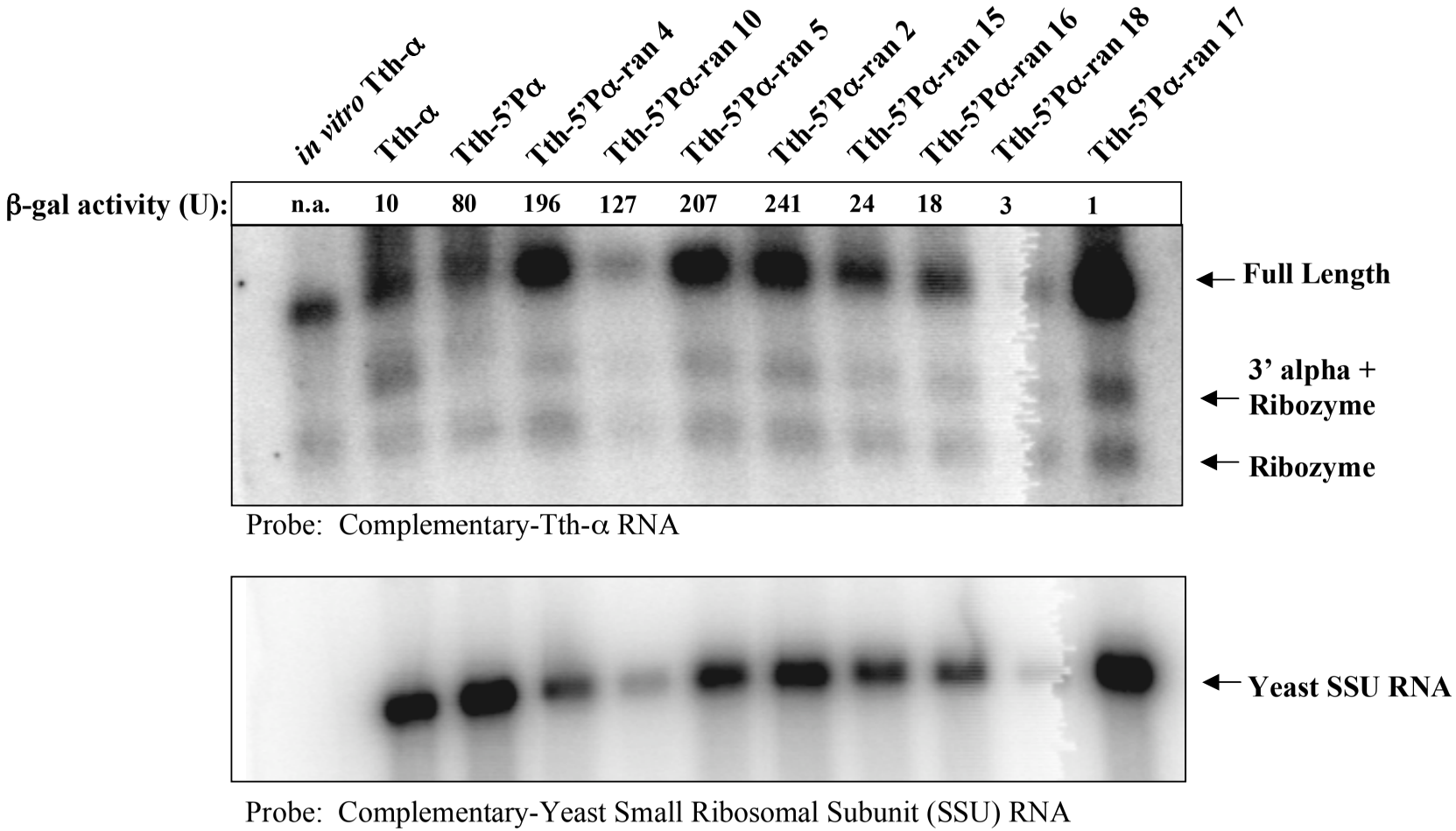
My examination of the folds predicted by these programs found that no common structures that correlate with high or low expressing sequences (data not shown). Indeed, several of the secondary structures predicted by RNAfold are shared between high and low expressing introns. There is also no correlation between expression levels and base pairing of the intron 5' end or the α start codon, two factors that could potentially allow or prevent access to the ribosome.

Since no sequence or structural motif could be linked to either high or low expression, I decided to examine the state of the intron RNA for clues that would explain the differences in expression between the introns. It seemed possible that the

new 5' end sequence might play a role in RNA stability, either by prevent processing or by inhibiting RNA degradation. I performed a northern blot on total RNA isolated from eight of the yeast strains containing Tth-5'P-α-ran introns (Figure 4.15). No differences were observed between the intron RNAs. All of the intron RNAs showed processing at IPS3 and at IPS1 or IPS2 (or both). Since experiments with Ppo.L1925 found that preventing IPS1 processing increased expression only four-fold (Lin and Vogt 1998), it seems unlikely that a lack of processing at IPS1 or IPS2 is the source of the high levels of expression seen for some of the Tth-5'P-α-ran introns. Differences in the levels of full-length spliced intron RNA also do not appear to be the source of differences the amount of α-fragment produced. Figure 4.16 depicts a comparison of the amount of full-length intron RNA, normalized by the amount of small ribosomal subunit RNA in each sample. These strains showed less than a four-fold difference in this number, with no correlation between the amount of full-length RNA and β-galactosidase activity. Thus, levels of full-length RNA and the extent of processing do not explain the differences in expression among these introns.

As an additional test, I examined if changing the Tth-α 5' end sequence, without the 5'P changes, could produce similar increases in β-galactosidase activity. Again, a randomization strategy was used to create Tth-α-ran, a pool of introns with randomized Tth.L1925-like 5' ends. The Tth.L1925 5' end is smaller than the 5'P sequence, with only four positions to randomize (Figure 4.17). Thus, the maximum number of possible introns in the pool was only 256. To make the Tth-α-ran pool, I followed a process similar to the one used to make the Tth-α-5'P pool. I picked 24 individual colonies, performed β-galactosidase assays on them, and sequenced the introns from eight (Figure 4.18). All 24 yeast strains exhibited levels of β-galactosidase activity similar to Tth-α, and none of the introns exhibited the high levels of expression seen from some of the Tth-5'P-α-ran introns. This result suggests

Figure 4.15: Northern blot of Tth-5'P-α-ran intron RNA. The top northern blot was probed with an *in vitro* transcribed radiolabeled RNA that is complementary to the Tth-α sequence. The bottom blot was probed with an *in vitro* transcribed radiolabeled RNA that is complementary to the yeast small ribosomal subunit sequence to serve as a loading control. Lane 1: 0.1 μg of in vitro transcribed Tth-α RNA. Lane 2-10: 2.5 μg of Total RNA isolated from a NOY505GalΩT strain with the listed intron integrated into the rDNA.


in vitro Tth-α
Tth-α
Tth-5'Pα
Tth-5'Pα-ran 4
Tth-5'Pα-ran 10
Tth-5'Pα-ran 5
Tth-5'Pα-ran 2
Tth-5'Pα-ran 15
Tth-5'Pα-ran 16
Tth-5'Pα-ran 18
Tth-5'Pα-ran 17
β-gal activity (U):
n.a.
10
80
196
127
207
241
24
18
3
1
Full Length
3' alpha +
Ribozyme
Ribozyme
Probe: Complementary-Tth-α RNA
Yeast SSU RNA
Probe: Complementary-Yeast Small Ribosomal Subunit (SSU) RNA

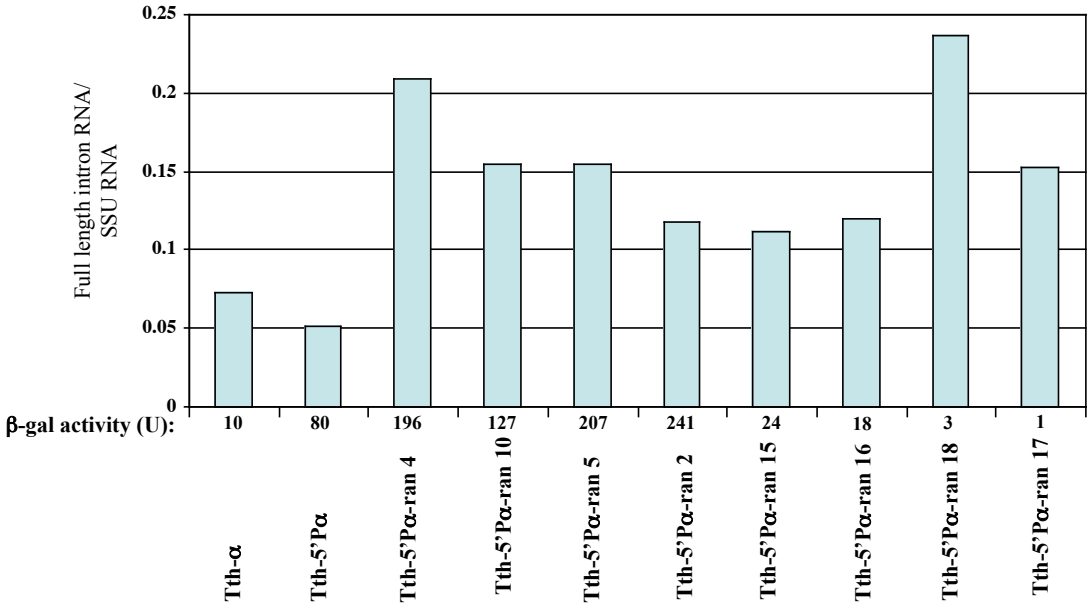
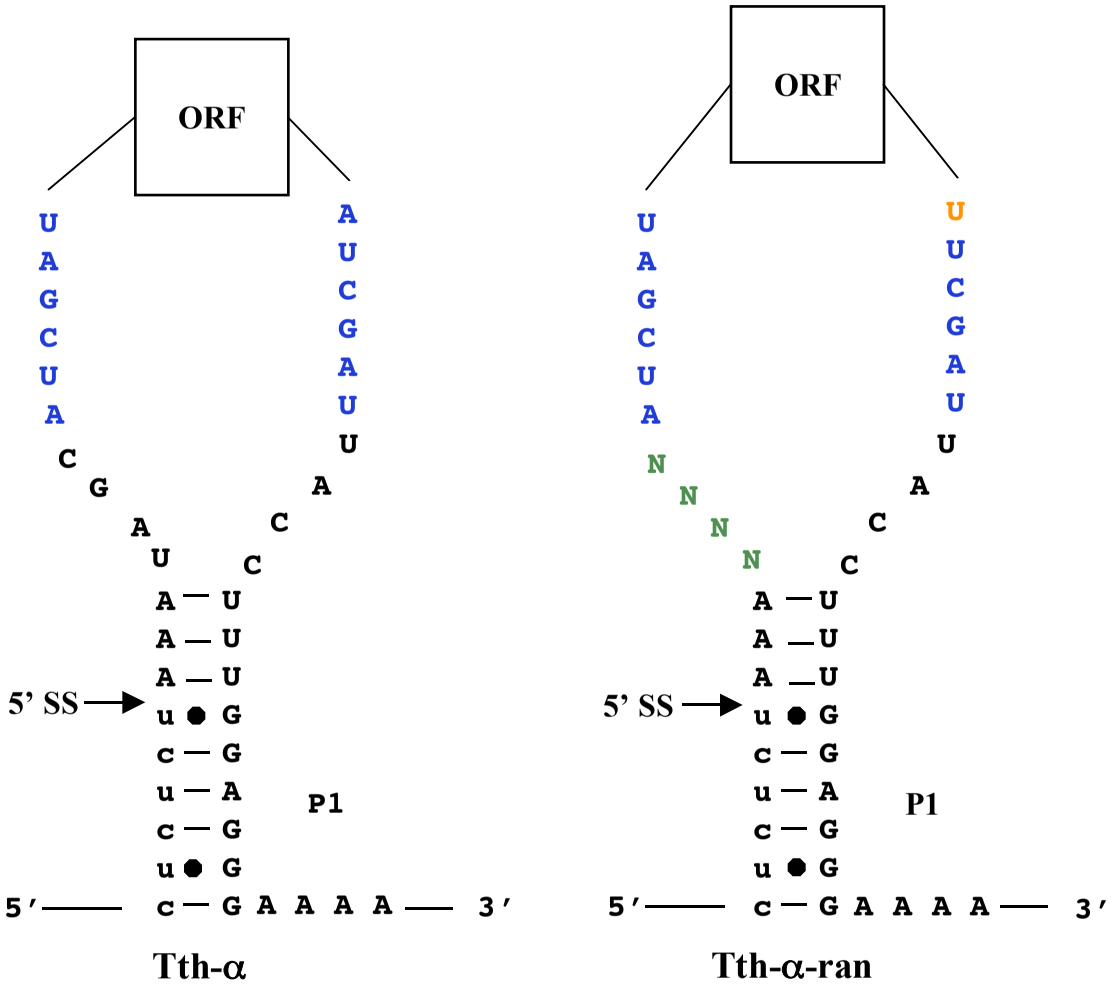
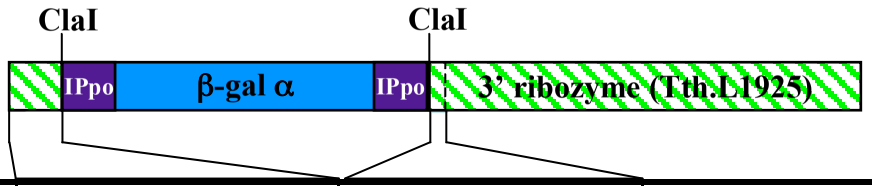


Figure 4.16: Full-length intron RNA levels normalized by comparison with SSU RNA levels. RNA levels were taken from figure 4.15. The β-galactosidase activity of each intron is noted below their respective column.

Figure 4.17: Randomization of the Tth-α 5' end. Nucleotides in green are randomized in the Tth-α-ran pool. The mutation that disrupts the 3' ClaI site is shown in orange. Nucleotides composing the ClaI sites are shown in blue. Intron sequence is in caps while the 5' exon sequence is in lower case.


ORF
U A G C U A C G A C A U C
A U C G A U U A C C
A — U
A — U
A — U
5' SS
u ● G
c — G
u — A
c — G
u ● G
P1
5' c — G A A A A — 3'
Tth-α
ORF
U A G C U A N N N N C
U U C G A U U A C C
A — U
A — U
A — U
5' SS
u ● G
c — G
u — A
c — G
u ● G
P1
5' c — G A A A A — 3'
Tth-α-ran



Intron	Sequence (5' end)	Sequence (P1')	β-galactosidase activity (U)
Tth-α	AAAUAGC	UACCUUUGGAGGG	10
Tth-α-ran 1	AAACACU	UACCUUUGGAGGG	16
Tth-α-ran 2	AAAAUAC	UACCUUUGGAGGG	5
Tth-α-ran 3	AAACAGA	UACCUUUGGAGGG	17
Tth-α-ran 4	AAACCCU	UACCUUUGGAGGG	19
Tth-α-ran 5	AAAAUCU	UACCUUUGGAGGG	13
Tth-α-ran 6	AAAGCUC	UACCUUUGGAGGG	11
Tth-α-ran 7	AAAUAGU	UACCUUUGGAGGG	7
Tth-α-ran 8	AAAACAG	UACCUUUGGAGGG	4

Figure 4.18: Tth-α-ran intron 5' end sequence and β-galactosidase activity. Red letters in the second column indicate the wild type Tth-α sequence and green letters in the second column indicate nucleotides in the randomized region. Red letters in the third column indicate the nucleotides that are not shared between Tth-α and Tth-5'P-α.

Intron	Sequence (5' end)	Sequence (P1')	β-galactosidase activity (U)
Tth-α-ran 10	Not sequenced	UACCUUUGGAGGG	8
Tth-α-ran 11	Not sequenced	UACCUUUGGAGGG	4
Tth-α-ran 12	Not sequenced	UACCUUUGGAGGG	14
Tth-α-ran 13	Not sequenced	UACCUUUGGAGGG	10
Tth-α-ran 14	Not sequenced	UACCUUUGGAGGG	11
Tth-α-ran 15	Not sequenced	UACCUUUGGAGGG	10
Tth-α-ran 16	Not sequenced	UACCUUUGGAGGG	14
Tth-α-ran 17	Not sequenced	UACCUUUGGAGGG	14
Tth-α-ran 18	Not sequenced	UACCUUUGGAGGG	8
Tth-α-ran 19	Not sequenced	UACCUUUGGAGGG	11
Tth-α-ran 20	Not sequenced	UACCUUUGGAGGG	8
Tth-α-ran 21	Not sequenced	UACCUUUGGAGGG	12
Tth-α-ran 22	Not sequenced	UACCUUUGGAGGG	13
Tth-α-ran 23	Not sequenced	UACCUUUGGAGGG	9
Tth-α-ran 24	Not sequenced	UACCUUUGGAGGG	4

Figure 4.18 (Continued): Tth-α-ran introns 10-24 were not sequenced. Red letters in the third column indicate the nucleotides that are not shared between Tth-α_and Tth-5'P-α.

that the 5'P sequence is able to modify expression of intron-encoded ORFs in a way that the native Tth-α 5' end is not, although how the 5'P sequence modifies translation remains unknown.

Discussion

Role of the 3'UTR

The results detailed in this chapter, summarized in Figure 4.19, show that intron sequences can influence expression levels from chimeric Tth.L1925 introns. The Ppo.L1925 3'UTR sequence has a constant effect on expression from chimeric Tth.L1925, as every intron made with a UTR sequence expressed at a lower level than the corresponding UTR-less intron. Expression from Tth-α is reduced slightly, with Tth-αUTR producing half the amount of β-galactosidase activity. The effect of the 3'UTR is more striking when it is added to the highly expressing Tth-5'P-α, which it reduces nearly twenty fold. Interestingly, the ability of the 3'UTR sequence to lower expression appears to be dominant over the ability of the 5' end to increase expression, as the Tth-αUTR and Tth-5'P-αUTR introns express at the same level.

The HEUTR sequences, which dramatically increase the expression in Ppo-α introns, have a similar inhibitory effect on Tth-α expression as the 3'UTR sequence. All of the three HEUTR sequences tested decrease levels of expression in both Tth-α and Tth-5'P-α introns. The fact that adding the 5'P sequence to the Tth-α-HEUTR introns does not increase expression suggests that, in Ppo-α, the 5'P sequence is not involved in the HETUR increase in expression or, if the 5'P sequence is involved, additional Ppo.L1925 sequences are required to increase expression.

Role of the 5' end

The intron 5' end sequence appears to have a large influence on levels of translation from Tth-α. Adding the *Physarum* 5'P sequence to Tth-α increases expression from the intron nearly ten-fold. This increase does not require the exact

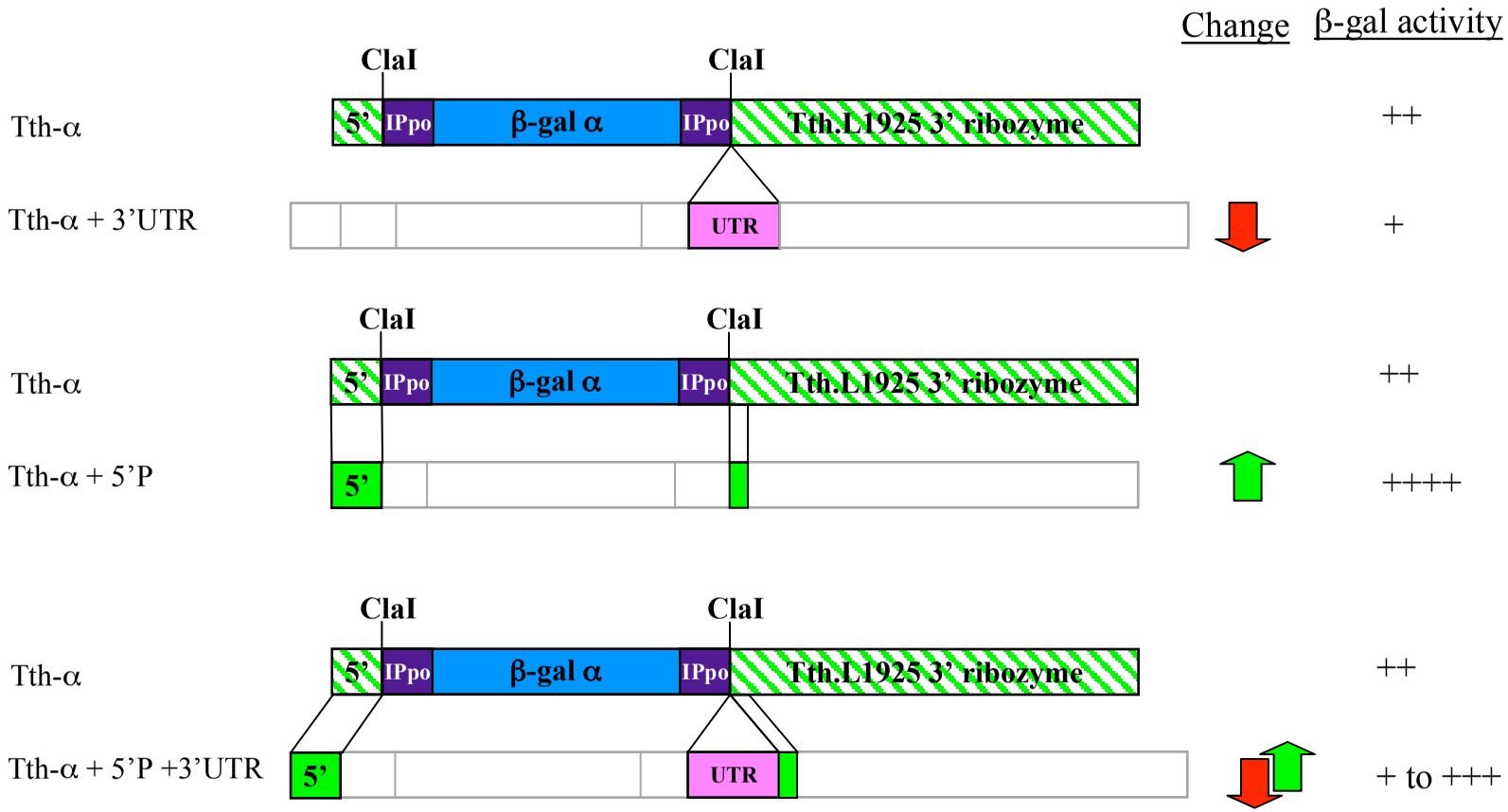


Figure 4.19: The effect of sequence changes on expression of the intron-encoded ORF. Red arrows indicate a decrease in expression, green arrows indicate an increase in expression, and N/C indicates no change in expression.

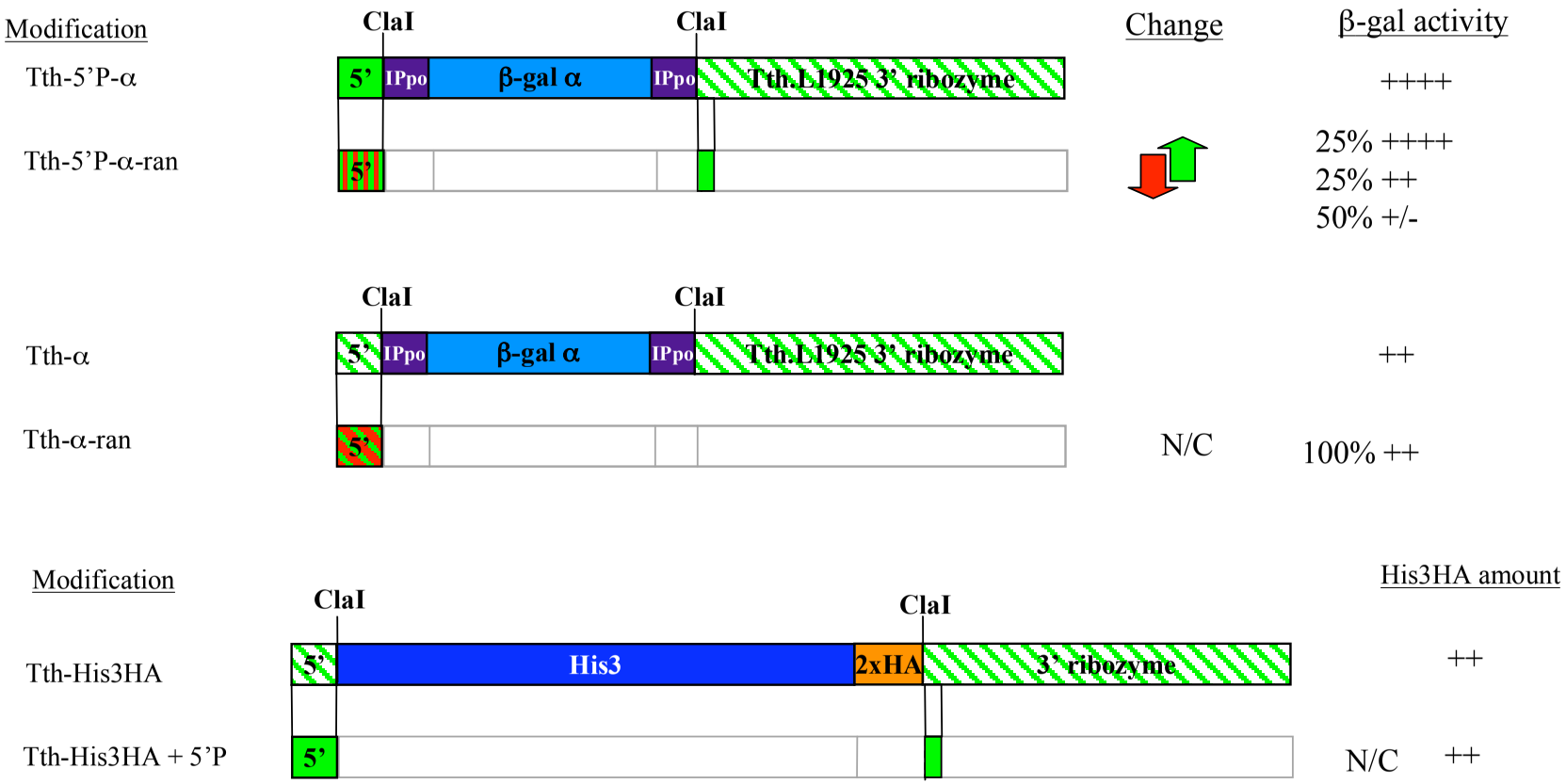


Figure 4.19 (continued): Red arrows indicate a decrease in expression, green arrows indicate an increase in expression, and N/C indicates no change in expression.

5'P sequence, as one quarter of the introns in the Tth-5'P-α-ran pool increased expression by a similar amount (or more). It is unclear how the 5' end sequences are influencing translation since no consensus sequence that increases expression is apparent. Although high expressing introns from the Tth-5'Pα-ran pool tend to be AT rich, the first four bases of the original Tth-5'Pα sequence is a string of Cs. Stop codons in the 5' end sequence do not diminish activity either. While the last two nucleotides of the 5' end sequence of Tth-5'α-ran 16 and Tth-5'α-ran 17, two of the weakly expressing introns, form a UAA stop codon when combined with the first nucleotide beyond the randomized sequence, high-expressing Tth-5'α-ran 1 contains the same UA sequence at positions 7 and 8. Two other high expressers, Tth-5'α-ran 2 and Tth-5'α-ran 9, contain UAA stop codons at other positions with no apparent reduction of translation. Both Tth-5'α-ran 16 and Tth-5'α-ran 17 contain out of frame start codons, so perhaps the low β-galactosidase activity from these introns could be explained by translation starting at these codons. This model does not explain, the low β-galactosidase activity of other Tth-5'α-ran introns that lack start codons in the randomized sequence.

In silico secondary structure prediction of high expressing and low expressing intron sequence produced similar contradictory results. There is no predicted secondary structure motif common to either high or low expressing introns. Since the ribosome needs to access the start codon to begin translation, I hypothesized that the predicted structures of high expressing introns would have α start codons that were not base paired. This is not the case, as the all of the predicted structures have at least one base of the start codon involved in a pair, and several of the high expressing sequences have all three bases paired. Structures of low expressing sequences also have a mixture of partially- and fully-paired start codons.

The hypothesis that interactions between the 5' end sequence and the α or ribozyme sequence allow the ribosome access to the α ORF is also not supported by the predicted structures. Individual intron 5' ends are unpaired, interact with other sequences close to the 5' end to forming a short stem loop, interact with sequences near the 3' end of the α ORF to extending the P1' stem, or interact with sequences at the 3' end of the intron. There is no correlation between any of the 5' end interactions and expression level of the intron.

It is also difficult to explain why all of the introns from the Tth-α-ran pool had levels of expression similar to Tth-α, while introns from the Tth-5'P-α-ran pool had varied levels of expression. Neither of the two obvious differences between the introns in these two pools, the slightly longer length of the Tth-5'P-α-ran 5' end and the differences in the P1 stem downstream of the ORF (Figure 4.9), provides an explanation for differences in activity. There is no correlation with 5' end length and expression levels, as individual introns from the Tth-5'P-α-ran pool, all of which have a longer 5' end than the Tth-α-ran pool introns, have both lower and higher expression levels than the Tth-α-ran introns. The same is true for the downstream P1 sequences possessed by all Tth-5'P-α-ran introns and lacked by the Tth-α-ran introns.

In contrast to its effect on Tth-5'Pα, adding the Ppo.L1925 5' end sequence had little effect on Tth-His3HA. The 5' end might need to interact with sequences in the α ORF, or the I-PpoI sequences at the 5' and 3' ends of the ORF, to increase expression. The length of the His3HA ORF could also prevent interactions between the 5' end and sequences in the ribozyme. Lastly, sequences in His3HA could interact with that the 5' end thereby preventing it from increasing translation. For example, His3HA nucleotides 64-84, 5'-UUUAAAGGGTG-3', could interact with the 5' end sequence of 5'-CACCCCCUUAA-3'.

Neither the 3'UTR nor 5' end sequence influenced processing of chimeric Tth-α introns. All of the Tth-α introns, including those that have no 3'UTR sequence, processed at the same three sites, IPS1, IPS2, and IPS3. Two of these sites, IPS1 and IPS2, match the processing sites of the same name in Ppo.L1925. The fact Tth.L1925 introns lacking the 3'UTR sequences still processed at these sites implies that the 3'UTR may not be involved in processing in Ppo.L1925, even though the 3' end of the 3'UTR is defined by processing.

In summary, intron sequences do play a large role in the levels of expression from chimeric Tth.L1925 introns. Adding the Ppo.L1925 3'UTR sequence to Tth-α or Tth-5'P-α lowered expression levels from both introns. None of the HEUTR sequences increased expression in the Tth-α, or Tth-5'P-α contexts as they do in Ppo-α. Instead, the HEUTR sequences also lowered expression levels from those introns. I found that the 5' end sequence, on the other hand, could greatly increase expression from Tth-α. Addition of the 5'P sequence to Tth-α increased α expression almost ten fold and the screen of Tth-5'P-α introns with randomized 5' end sequences showed that one quarter of 5'P-like 5' ends exhibited this high level of expression. Randomization of the Tth-α 5' end, without the 5'P sequence, showed that none of the possible Tth.L1925-like 5' end sequences could increase expression in a similar way. It is unclear how these sequences are modifying expression, although they could be promoting the formation of a highly-translated intron species or interactions with yeast factors that promote translation.

Chapter V

Investigation of the Mechanism of Translation of Chimeric *Tetrahymena* Intron-Encoded ORFs

Introduction

Changing the sequence 5' and 3' of an intron-encoded ORF can increase or decrease the expression of a Ppo.L1925- or Tth.L1925-encoded ORF. The mechanism by which these sequences act, however, is unclear. High-expressing intron sequences do not appear to affect protein levels by changing the rate of splicing, by changing the position of processing sites, by altering the levels of full-length intron RNA, or by stabilizing the intron RNA molecule. Alternative models explaining the effect of these sequences include promoting interaction of the intron RNA with cellular translation machinery or promoting the formation of an intron structure necessary for translation.

One model for translation of intron-encoded ORFs is that they use internal ribosome entry sites (IRES), which promote translation in a cap-independent manner. This model is attractive for intron RNA translation as spliced introns are also cap-less. At least two well-characterized IRES elements have been documented in yeast, the *URE2* IRES and the invasive growth IRES. The *URE2* IRES forms a 25 nucleotide long stem topped with a 24-nucleotide loop (Reineke, Komar et al. 2008). Since this 74-nucleotide sequence is much longer than the intron 5' end, it is unlikely that the intron sequence is mimicking the *URE2* IRES. The invasive growth IRES, on the other hand, could be mimicked by the intron 5' end sequences, as both elements are characterized by short (~12 nucleotide) unstructured poly-A stretches directly upstream of the respective ORF start codon (Figure 5.1 A) (Gilbert, Zhou et al. 2007), and many of the high-expressing introns from the Tth-5'P-α-ran pool had A-rich 5' end sequences.

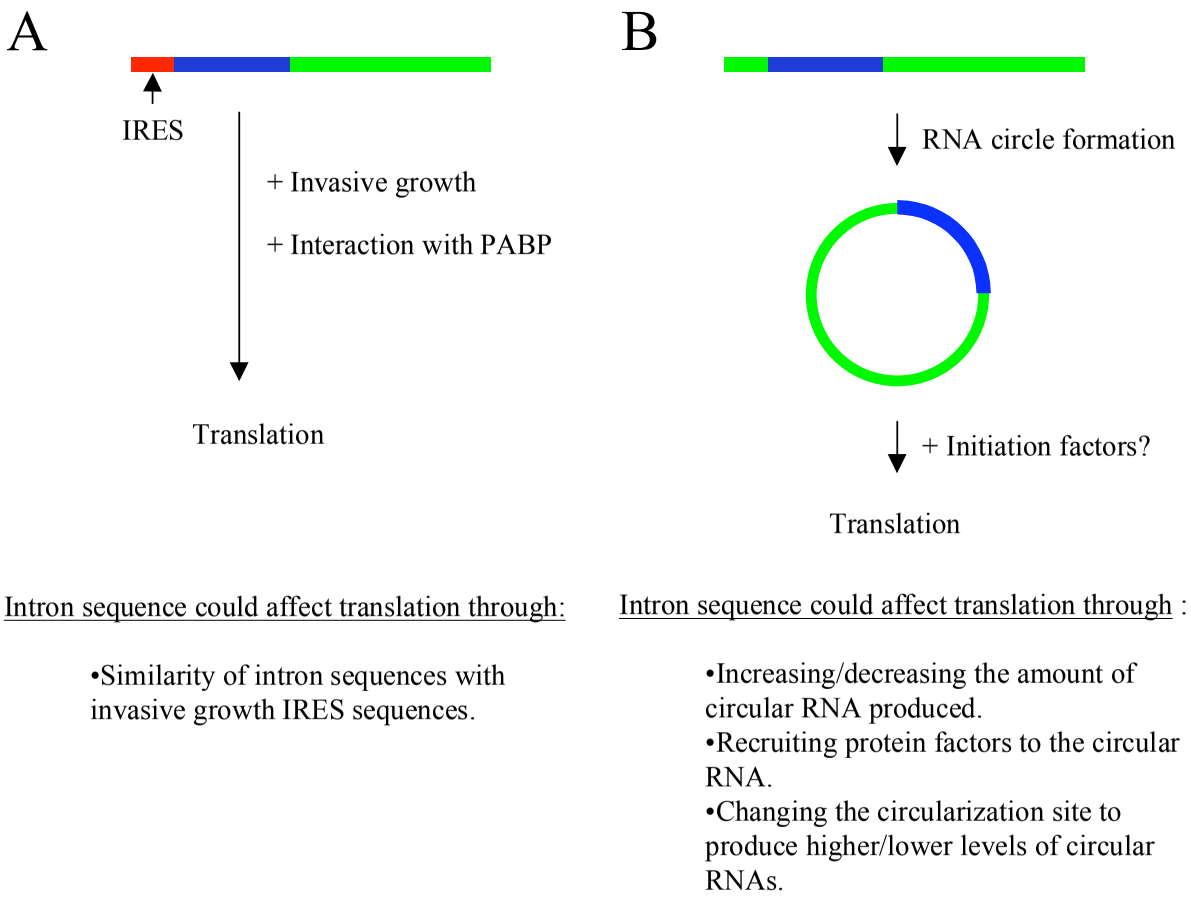


Figure 5.1: Two models for translation of intron-encoded ORFs. A. Interaction with the yeast invasive growth IRES pathway. The intron RNA interacts with factors, including poly-A binding protein, that allow expression from IRES sequences in invasively growing yeast. B. Intron circularization leads to translation. The intron forms a circular RNA, which allows interaction with protein factors with which the linear intron RNA does not. Interact.

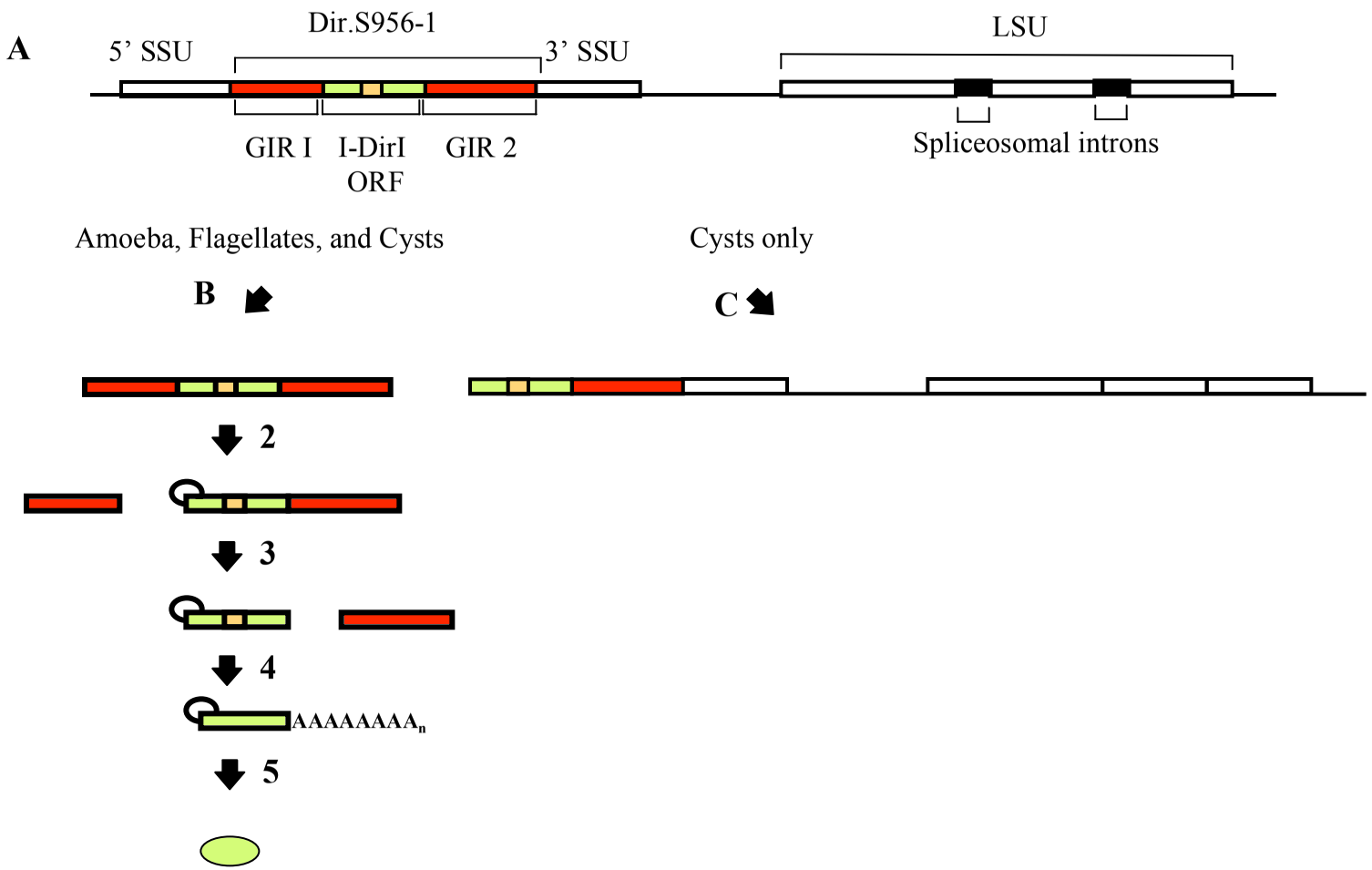
It is also possible that the full-length intron RNA must undergo a post-splicing reaction in order to serve as a viable message (Figure 5.1 B). Cleavage at internal processing sites does not appear to promote translation, but the role of full-length intron circles has not been examined (Lin and Vogt 1998). These RNA circles might mimic the circular complexes that mRNAs form through interactions between the cap binding complex and poly-A binding proteins. The high expressing introns could be forming intron circles at higher levels or at different junctions to increase translation levels.

Results

Translation levels of Tth.L1925-encoded ORFs are constant throughout the yeast cell cycle.

Before investigating specific models, I examined if expression from Tth-α and Ppo-α changed during growth. Several instances have been described in which the population of intron RNAs changes depending on the growth state of the cell. For example, Dir.S951-1, an intron naturally found in the slime mold *Didymium iridis*, produces an additional RNA species in encysted cells that is not found in amoeboid or flagellate cells (Figure 5.2) (Vader, Johansen et al. 2002). In *Didymium*, the freshly-transcribed, unprocessed rRNA (the pre-rRNA) encodes both the small ribosomal subunit RNA (SSU) and the large ribosomal subunit rRNA (LSU). In growing *Didymium* cells, which can be amoebae or flagellates, Dir.S951-1 catalyzes its removal from the pre-RNA and then proceeds to go through a series of processing steps. In encysted *Didymium*, however, Dir.S951-1 processes internally before splicing itself out of the pre-rRNA; this results in the accumulation of a 7.5 kb RNA that includes the 3' half of Dir.S951-1, the SSU rRNA, the LSU rRNA, and a spacer rRNA that lies between the SSU and LSU. The biological role of this 7.5 kb RNA is

Figure 5.2: Splicing pathways of Dir.S956-1 vary depending on cellular growth stage. A. The *Didymium* pre-rRNA. B. In *Didymium* amoeba, flagellates, and cysts, GIR2 splices the Dir.S956-1 intron RNA out of the surrounding RNA, and the intron RNA is processed by GIR1, which is responsible for the 5' cleavage, and cellular factors, which are responsible for the 3' cleavage. C. In encysted *Didymium*, there is an alternate processing pathway. GIR1 cleaves 5' of the intron-encoded ORF and the two spliceosomal introns are removed from the LSU to create a 7.5 kb RNA.


A
5’ SSU
Dir.S956-1
3’ SSU
LSU
GIR I
I-DirI ORF
GIR 2
Spliceosomal introns
Amoeba, Flagellates, and Cysts
Cysts only
B
C
2
3
4
AAAAAAAA$_n$
5

unclear, but since it does contain the homing endonuclease ORF, it could act as a source of homing endonuclease for the encysted cell.

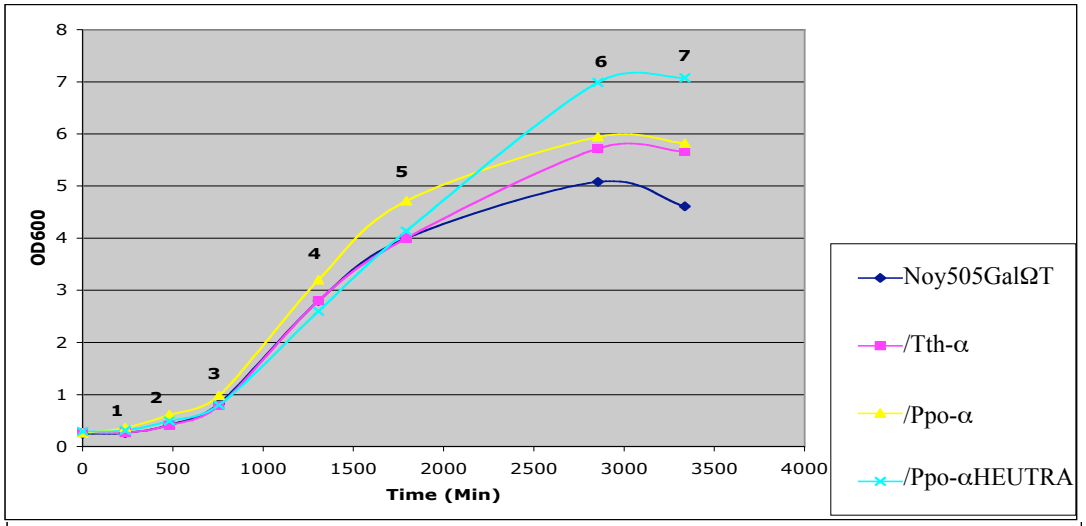
I decided to examine if expression of intron-encoded ORFs is influenced by the growth stage of their host yeast cell. I grew 200 ml cultures of NOY505GalΩT yeast with no intron, Tth-α, Ppo-α, and Ppo-α-HEUTR A, and then periodically performed β-galactosidase assays (Figure 5.3). All of the yeast strains had comparable doubling times and β-galactosidase levels were consistent throughout all stages of growth. These results rule out the possibility that growth-stage specific factors are affecting the translation of the intron-encoded ORF.

The *YMR181c* IRES sequence does not increase expression of Tth.L1925-encoded ORFs.

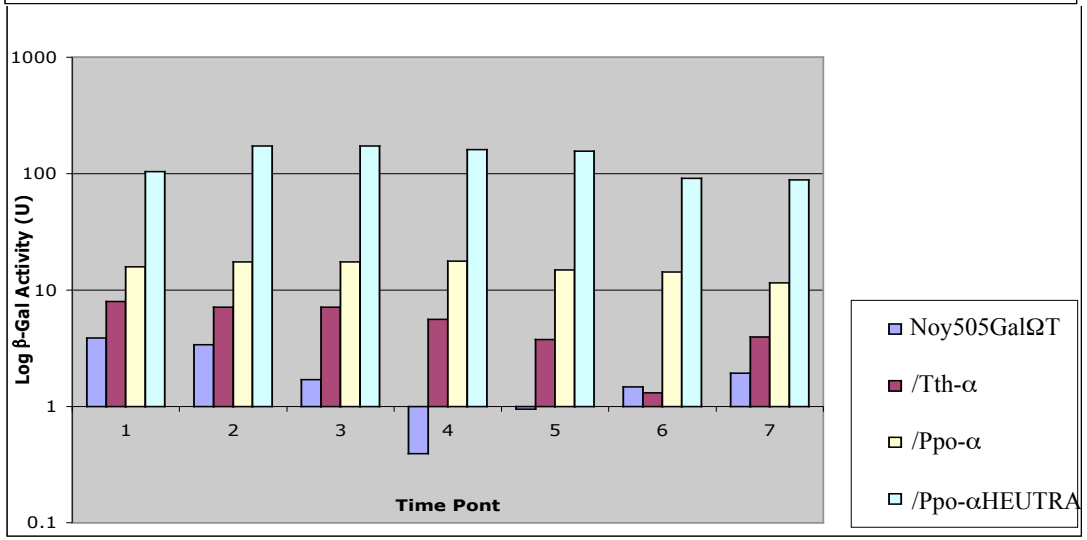
S. cerevisiae cells have two growth forms, a vegetative form and an invasive (or filamentous) form (Gimeno, Ljungdahl et al. 1992). The vegetative form is characterized by round cells, while the invasive form is characterized by elongated cells that clump together in liquid culture and penetrate the surface of agar plates. The carbon source available to the yeast determines which of these forms the cells take; yeast grows vegetatively on glucose-containing media and invasively on media with other carbon sources, such as galactose (Palecek, Parikh et al. 2002). Yeast cells that have been grown on glucose and then switched to glucose-lacking media respond by undergoing a number of sudden changes. Translation stops immediately (within 1 minute) through a 5' decapping enzyme-dependent process, which likely involves the decapping of the majority of mRNA, and the cells then transition to invasive growth, which requires the cells to undergo a number of structural changes (Ashe, De Long et al. 2000; Cullen and Sprague 2000). An inherent contradiction exists between these two responses, as translation is required to produce the factors necessary for the switch to invasive growth. Yeast has solved this conflict through the invasive growth IRES

Figure 5.3: The β-galactosidase activity of intron-containing yeast strains through out growth. A. Growth curve of Noy505GalWT, Noy505GalWT/Tth-α, Noy505GalWT/Ppo-α, and Noy505GalWT/Ppo-αHEUTR A. Cells were grown in YEPG media shaking at 30˚C. Aliquots were taken at time points 1-7 and tested for b-galactosidase activity (shown in B).

A


OD600
Time (Min)
0
500
1000
1500
2000
2500
3000
3500
4000
1
2
3
4
5
6
7
Noy505GalΩT
/Tth-α
/Ppo-α
/Ppo-αHEUTRA

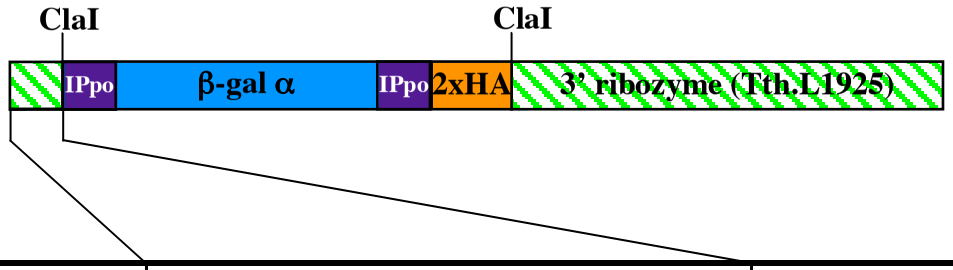
B


Log β-Gal Activity (U)
Time Pont
1000
100
10
1
0.1
1
2
3
4
5
6
7
Noy505GalΩT
/Tth-α
/Ppo-α
/Ppo-αHEUTRA

sequence. This sequence, found upstream of the genes required for invasive growth, allows for translation of these genes during the starvation-induced translation shutdown (Gilbert, Zhou et al. 2007). These IRES sequences have been shown to operate in a poly-A binding protein (PABP)-dependent manner and likely initiate translation by recruiting PABP, which in turn recruits initiation factors or the 60S ribosomal subunit independently of the 5' cap binding protein.

One possible model for expression of intron-encoded ORFs is that that capless intron RNA acts like an invasive growth gene that has an IRES. Higher expressing introns could have sequences better at mimicking the invasive growth IRES than lower expressing introns. To address this possibility, I made a chimeric Tth-α intron containing the invasive growth IRES from *YMR181c*, an invasive growth gene that is translated from the second message of a bicistronic mRNA (He, Li et al. 2003; Suzuki, Hori et al. 2003). The *YMR181c* IRES sequence contains an unstructured poly-A tract, AACAAAAAAAAA, directly 5' of the start codon. This sequence has been shown to act as an IRES both *in vitro* and *in vivo* , and deleting this sequence or replacing it with its reverse complement, UUUUUUUUUGUU, abrogates IRES activity (Gilbert, Zhou et al. 2007). If the high-expressing introns are mimicking the invasive growth IRES sequences, a Tth-α intron containing the IRES sequence should express at a higher level than an intron without it.

To determine if the IRES sequence has an effect on expression from Tth-α, I constructed three introns: Tth-αHA, Tth-IRES-αHA and Tth-RCIRES-alpahHA (Figure 5.4). Tth-αHA has an unmodified (Tth-α-like) 5' end, Tth-IRES-αHA has the *YMR181c* IRES sequence directly 5' of the α start codon, and Tth-RCIRES-αHA has the reverse complement of the *YMR181c* IRES sequence. Since it could be useful to track expression of the intron-encoded ORF without relying on expression of the Ω



Intron	Sequence (5' end)	β-galactosidase activity (U)
Tth-αHA	**AAAUAGCAUCGAUAUG**	7
Tth-IRESαHA	**AAAUAGCAUCGAUAACAAAAAAAAAAUG**	16
Tth-RCIRESαHA	**AAAUAGCAUCGAUUUUUUUUUUGUUAUG**	20

Figure 5.4: Effect of the *YMR181c*_IRES sequence on translation of intron-encoded ORFs. The ClaI site is in blue, the *YMR181c* IRES and the reverse complement of the YMR181c IRES sequences are in red and the start codon is in green.

fragment, I added a double HA tag to the C-terminus of the α sequence. The Tth-αHA intron serves as a control to ensure that the HA tag sequence does not influence expression or splicing. Unfortunately, attempts to visualize the αHA product through western blot were unsuccessful, possibly due to the small size of the α fragment (data not shown). However, after trans-integrating these introns into NOY505GalΩT, I found the β-galactosidase activity of Tth-αHA to be 7 U. This level is similar to that of Tth-α, which suggests that the HA tag sequence does not interfere with splicing, translation, or α-complementation.

Since yeast cells with galactose as the only carbon source grow invasively, the IRES sequences should be active in yeast growing in the media necessary to induce Ω fragment expression from NOY505GalΩT. If the Tth-α introns mimic invasive growth IRESs, Tth-IRES-α should exhibit higher β-galactosidase activity than Tth-RCIRES-α in invasively growing yeast. However, I found that in fact both Tth-IRES-αHA and Tth-RCIRES-αHA exhibit similar levels of β-galactosidase activity, 16 U and 20 U, respectively. This activity is slightly higher than Tth-αHA but is not at the extreme levels seen for the Tth-5'P-α-ran or Ppo-α-ran introns. The similarity between the levels of β-galactosidase activity of these two introns suggests that Tth-α intron RNAs are not translated by mimicking invasive growth IRES sequences.

Chimeric Tth.L1925 introns form several species of RNA circle

A second model for translation of intron-encoded ORFs involves the circularization of the intron RNA. During translation initiation of an mRNA, the poly-A binding proteins associate with the cap-binding complex, causing the mRNA to form a circle-like structure (Imataka, Gradi et al. 1998; Kessler and Sachs 1998). It could be that circular intron RNA is translated by somehow mimicking the pre-initiation mRNA. Intron 5' end and 3'UTR sequences could affect translation both by influencing the rate of circle formation or the junction at which the circles form. This

model also allows for 3'UTR sequences to influence translation by recruiting initiation factors, since in a circular RNA the 3'UTR would be upstream (as well as downstream) of the intron-encoded ORF.

Group I introns form two major species of RNA, full-length and internal guide sequence (IGS) circles (Zaug, Grabowski et al. 1983; Been and Cech 1986; Been and Cech 1987). Full-length circles are the result of an attack of the intron 3' end between the exogenous 5' G and the first nucleotide in the intron sequence resulting in a circle that contains the entirety of the intron sequence (Been and Cech 1986; Inoue, Sullivan et al. 1986; Nielsen, Fiskaa et al. 2003). This circle can re-linearize through a repeat of the first step of splicing; the intron ribozyme binds an exogenous G, which attacks at the circle junction between the 5' and 3' ends of the intron. Since the full intron sequence is preserved in full-length introns circles, a single spliced intron RNA could go through multiple circularization and un-circularization events. The second type of intron circle, an IGS circle, is the result of an attack of the intron 3' end in, or near, the IGS (Zaug, Grabowski et al. 1983; Been and Cech 1986; Been and Cech 1987).

Tth.L1925 has been shown to form full-length circles *in vitro*, but not *in vivo,* as well as two distinct IGS circles both *in vitro* and *in vivo* (Figure 5.5) (Been and Cech 1986; Been and Cech 1987; Nielsen, Fiskaa et al. 2003). The major Tth.L1925 IGS circularization product results from of an attack between nucleotides 14 and 15, while a minor product results from an attack between nucleotides 19 and 20. The circle species formed by Ppo.L1925 have not been as thoroughly studied as those of Tth.L1925, but Ppo.L1925 is known to produce full-length circles *in vitro* (Nielsen, Fiskaa et al. 2003). It is logical to expect that chimeric Tth.L1925 introns will form circles *in vivo*, although it is unclear if they form Tth.L1925-like IGS circles or Ppo.L1925-like full-length circles.

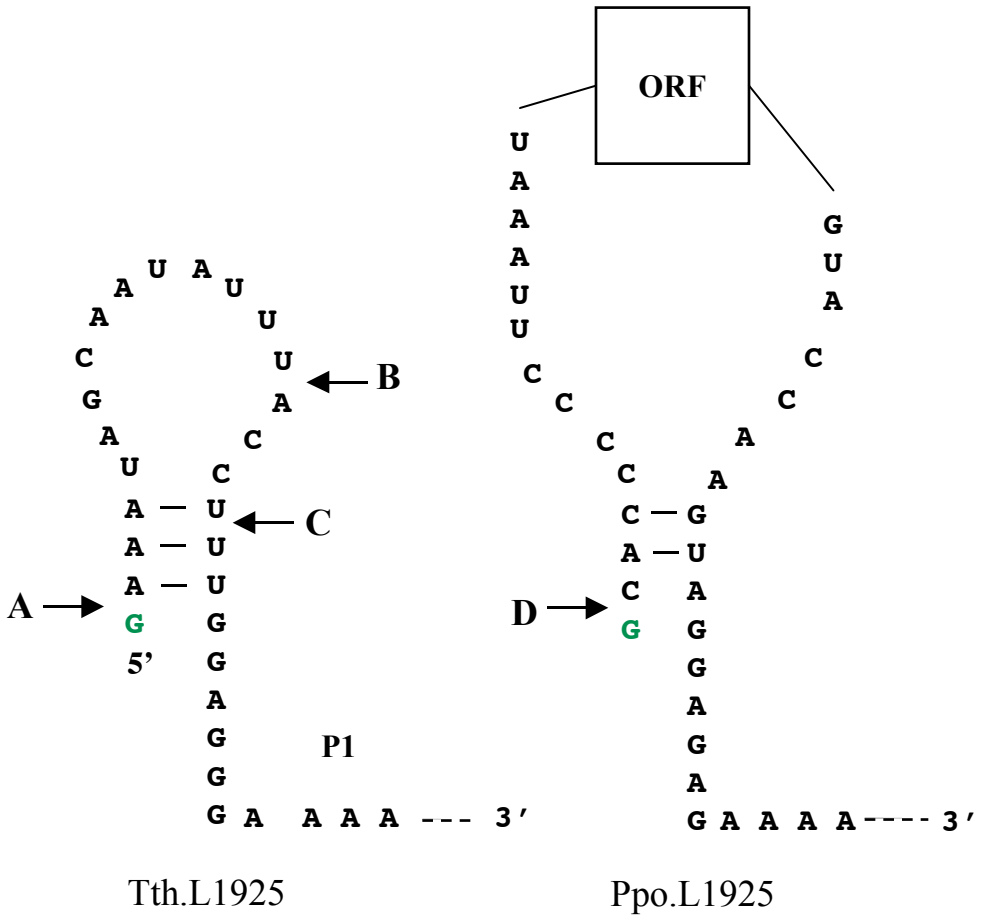


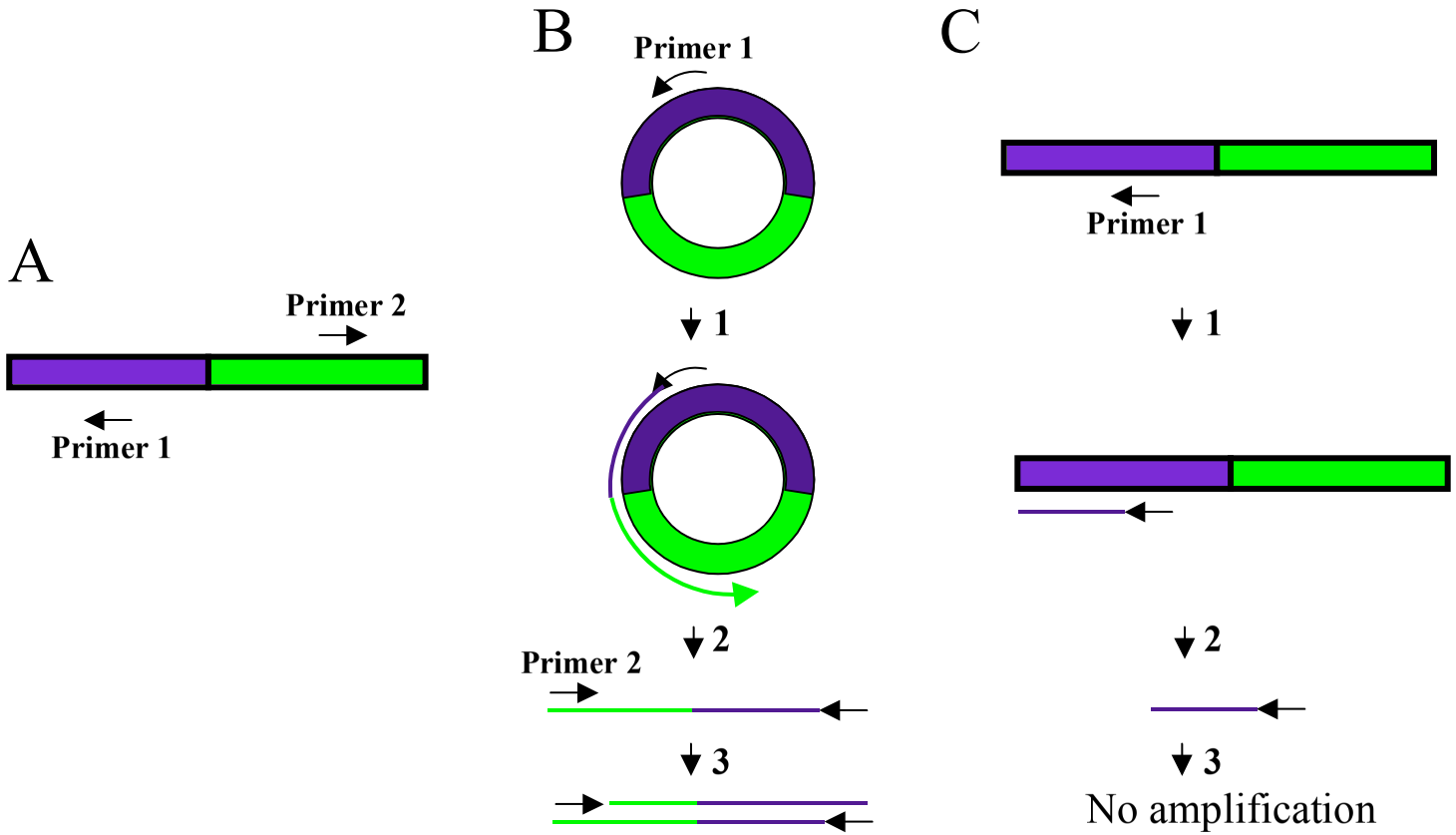
Figure 5.5: Circularization sites of Tth.L1925 and Ppo.L1925. The exogenous G added during the first step of splicing is shown in green. Circularization at A and D produce full-length circles, while circularization at B and C produces IGS circles. Tth.L1925 has been found to form circles at A only in vitro, while it forms circles at B and C both in vitro and in vivo. Circularization at B is the major species formed, while circularization at C is a minor product. Ppo.L1925 has been found to form circles at D only *in vitro*. The species formed by Ppo.L1925 *in vivo* have not been examined.

In order to determine if the chimeric Tth.L1925 introns circularize in yeast, and if there is a relationship between the circle species and intron-encoded ORF expression levels, I used reverse transcriptase-PCR (RT-PCR) to amplify RNA sequences containing circle junctions (Figure 5.6). I used two primers, one that binds to the 5' end of the intron (or the α sequence) and faces upstream and one that binds to the 3' end of the intron that faces downstream. Because of this primer arrangement, the only sequence amplified would be the result of intron circularization.

The initial primer set consisted of a primer that binds in the α sequence, 101 nucleotides from the 5' end of Tth-α, and one that binds 72 nucleotides from the 3' end of the ribozyme. I performed reverse transcription reactions on total yeast RNA isolated from NOY505GalΩT integrated with Ppo-α, Ppo-α-HEUTR A, Tth-α, Tth-5'P-α, and seven of the Tth-5'P-α-ran introns (Tth-5'P-α-ran 2, 4, 5, 10, 15, 16, and 17). If the intron RNAs formed full-length circles, the amplified product would be ~200 bp in length, and if the introns formed smaller circles, the amplified product would be correspondingly smaller. All amplifications from intron-integrated yeast RNA tested produced at least one band of the expected size (Figure 5.7), and several produced two. I infer that these bands result from amplification of the reverse-transcribed intron RNA, as RNA isolated from the intron-less parent strain, NOY505GalΩT (Figure 5.7 B, lanes 2 and 5), and PCR amplifications run from mock reverse transcriptions reactions did not produce bands (data not shown).

To determine the location of the circle junction, I sequenced a selection of the amplified PCR products. All of the sequences appeared to be amplified from a mixture of junctions. The most common junctions of these introns are listed in Figure 5.8. Ppo-α and Ppo-α-HEUTR A circularize at the same position, between intron nucleotides 5 and 6. Tth-α makes full-length circles, while the Tth-α-5'P circle junction is at the –8 position. High expressing Tth-α-ran 3 formed full-length circles,

Figure 5.6: Detecting circular intron RNA. A. Primer binding sites on the target intron RNA. The 5' half of the RNA is depicted in purple, the 3' half is depicted in green. B. Reverse-transcription and PRC over circular intron junction. 1. Reverse transcription starts at Primer 1 and transcribes over the 5' RNA, circle junction, and the 3' RNA. 2-3. PCR amplifies over the circle junctions. C. Reverse transcription and PCR of linear RNA will produce no product. 1. Reverse transcription starts at Primer 1 and transcribes over the 5' RNA, ending at the 5' end of the 5' end RNA. 2-3. PCR is unable to amplify the reverse transcription product, as there is no binding site for primer 2.


A
Primer 2
Primer 1
B
Primer 1
1
2
Primer 2
3
C
Primer 1
1
2
3
No amplification

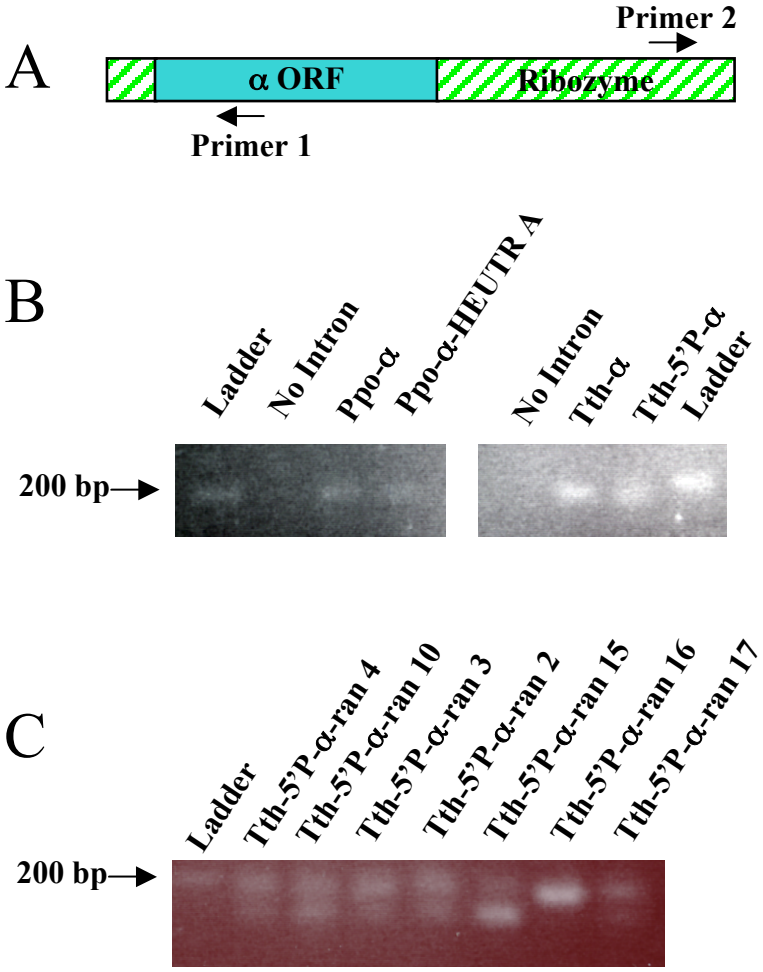


Figure 5.7: RT-PCR to detect circular RNA junctions near the intron 5' end. A. Primer binding sites for the reverse transcription-PCR. Primer 1 binds 101 nucleotides from the 5' end of a and primer 2 binds 72 nucleotides from the 3' ribozyme end. B. and C. Agarose gel electrophoresis of reverse transcriptase-PCR reactions of total RNA from NOY505GalWT with the indicated intron integrated

Intron	Circle Junction	β-gal Activity (U)
Ppo-α	-5	15
Ppo-αHEUTR A	-5	400
Tth-α	0 (Full Length)	10
Tth-5'P-α	-8	80
Tth-5'P-α-ran 3	0 (Full Length)	200
Tth-5'P-α-ran 4	-36 (α -20)	190
Tth-5'P-α-ran 15	-4	25
Tth-5'P-α-ran 16	-36 (α -20)	20

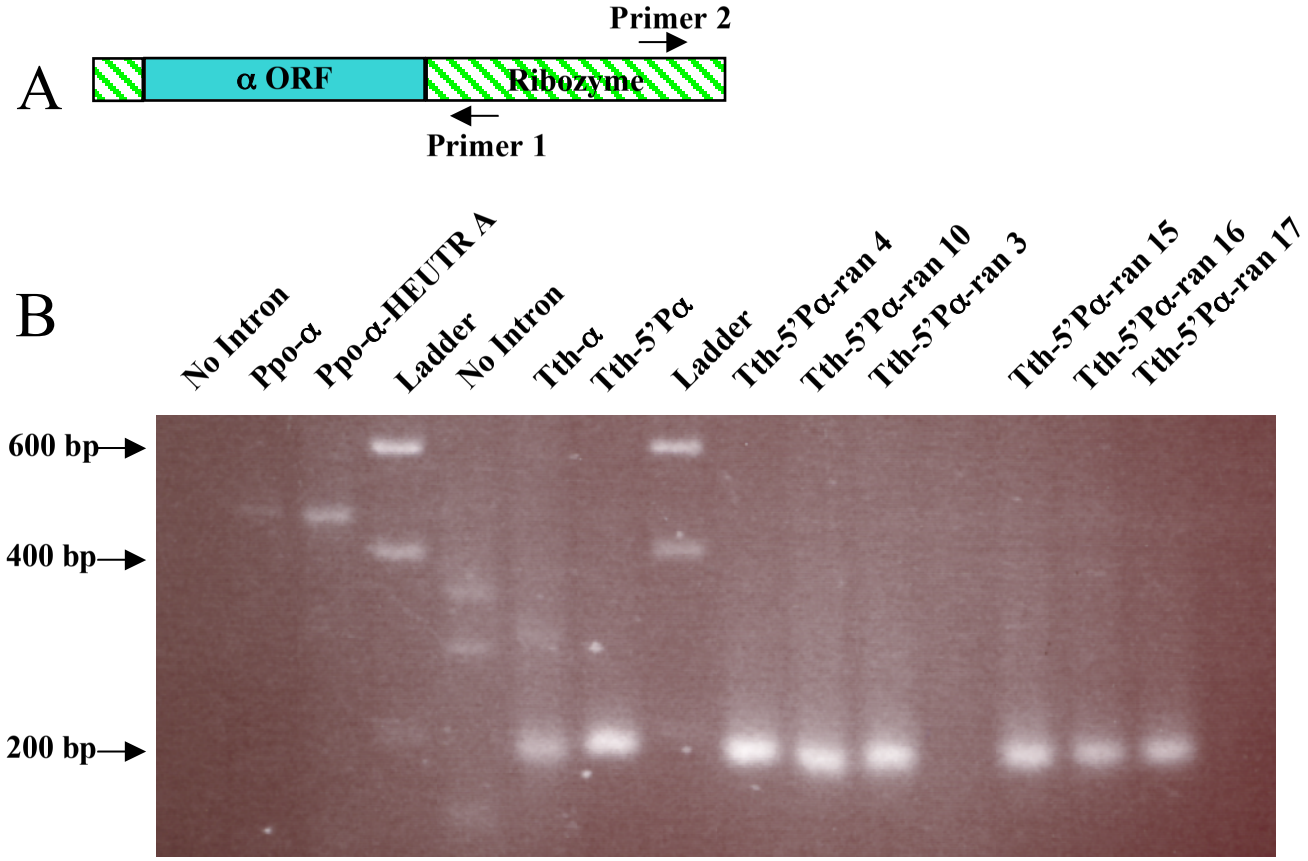
Figure 5.8: Circle junction and β-galactosidase activity of selected intron circles from figure 5.7.

while low expressing Tth-α-ran 15 circularized at the –15 position. Both high expressing Tth-α-ran 4 and low expressing Tth-α-ran 16 circularized at the same junction, -36. Since the –36 junction is beyond the β-galactosidase α start codon, and there is no in-frame start codon in the Tth.L1925 3' end, it is highly unlikely that these circles act as the message for the α protein.

Through examination of the raw data from the DNA sequencing reactions of the Tth-α circle, I discovered a minor circularization junction is at –92, which is the binding site of primer PrRS71. In order to guard against the possibility that primer binding during the reverse transcription reaction is somehow influencing circle formation, I performed another RT-PCR reaction using a primer that binds in the P2 stem, downstream of the α ORF (PrRS157 for Tth.L1925, PrRS158 for Ppo.L1925, along with either PrRS72 or PrRS71, respectively, for the subsequent amplification). Amplification of full-length circles from this primer should result in products ~500 nucleotides in length. Bands of this size were seen for both Ppo-α and Ppo-α-HEUR A (Figure 5.9 B). All Tth.L1925-based introns tested produced a much smaller product, ~200 bp in length. Sequencing of these bands showed that the junction at position 289 of the α ORF (the α ORF is 297 nucleotides long) (Figure 5.10). Formation of these circles cannot contribute to production of the β-galactosidase α protein, as the circular RNA contains almost none of the β-galactosidase α sequence and the 5' liberated sequence is almost certainly quickly degraded.

Since there is no apparent relationship between circle junctions formed and the amount of β-galactosidase activity produced by each intron-containing yeast strain (Figure 5.10), I concluded that that circle formation is unlikely to be involved in expression of the intron-encoded ORFs.

Figure 5.9: RT-PCR to detect circular RNA junctions near the P1' stem. A. Primer binding sites for the reverse transcription-PCR. Primer 1 binds 407 nucleotides from the 5' end of α and primer 2 binds 72 nucleotides from the 3' ribozyme end. B. and C. Agarose gel electrophoresis of reverse transcriptase-PCR reactions of total RNA from NOY505GalΩT with the indicated intron integrated.


A
Primer 2
α ORF
Ribozyme
Primer 1
B
No Intron
Ppo-α
Ppo-α-HEUTR A
Ladder
No Intron
Tth-α
Tth-5'Pα
Ladder
Tth-5'Pα-ran 4
Tth-5'Pα-ran 10
Tth-5'Pα-ran 3
Tth-5'Pα-ran 15
Tth-5'Pα-ran 16
Tth-5'Pα-ran 17
600 bp
400 bp
200 bp

Intron	Circle Junction	β-Gal Activity (U)
Ppo-α	-5	15
Ppo-αHEUTR A	-5	400
Tth-α	-302 (α -289)	10
Tth-5'P-α	-306 (α -289)	80
Tth-5'P-α-ran 3	-306 (α -289)	200
Tth-5'P-α-ran 4	-306 (α -289)	190
Tth-5'P-α-ran 15	-306 (α -289)	25
Tth-5'P-α-ran 16	-306 (α -289)	20

Figure 5.10: Circle junction and β-galactosidase activity of selected intron circles from figure 5.9. Numbers in parenthesis indicate a position in the α ORF.

Discussion

Capless, Pol I-derived RNA has been shown to be inefficiently translated in yeast, presumably by ribosomes aberrantly initiating translation on an abundant RNA at a low rate (Lo, Huang et al. 1998). Translation of intron-encoded ORFs may be more complicated than this model suggests, since intron sequences can greatly change the amount of protein being translated without changing the level of spliced intron RNA. In this chapter, I investigated two possible models to explain how intron sequences could modulate intron-encoded ORF expression in yeast. According to one model, the 5' part of the intron RNA acts like an invasive growth IRES, and according to the other intron circularization creates the messenger RNA for the intron-encoded ORF.

Invasive growth IRES element does not increase expression from chimeric Tth-α

The invasive growth IRES sequence allows yeast cells to switch from vegetative growth to invasive growth while at the same time conserving energy by stopping all other translation. The genes responsible for invasive growth escape starvation-induced translation repression because their messages possess an A-rich IRES sequence. Since many of the highly expressing Tth-5'Pα-ran introns also had A-rich 5' end sequences, it seemed possible that the introns were using this mechanism to initiate translation as well. However, the results did not support this model, as adding the *YMR181c* IRES sequence to the intron did not allow the intron-encoded ORF to be translated through this pathway.

Why is the ORF not translated from an IRES-containing intron RNA through the invasive growth IRES system? One difference between the intron RNA and invasive growth gene RNA is the length of sequence 5' of their ORFs. Many of the invasive growth IRES elements have long stretches of upstream sequence, and some, such as *YMR181c*, even have entire ORFs upstream. The small, shared A-rich

sequence is required for translation, as deletion of these sequences abrogates IRES function, but it is possible that a length of additional upstream sequence, which is lacked by the intron, is necessary for binding of other factors to facilitate IRES-mediated.

Chemical probing of the poly-A tract in the IRES element has revealed that the region is unstructured (Gilbert, Zhou et al. 2007). It is possible that the intron structure prevents PABP binding to the intron RNA. The P1 stem is only 10 nucleotides upstream of the poly-A tract, and the ClaI sequence directly 5' of the poly-A tract and 3' of the α ORF are likely to be involved in base pairing as well. According to this argument, the intron 5' end is far more structured than the IRES element RNA, which could prevent translation initiation.

Alternatively, the invasive growth IRES elements may only be active during the switch from vegetative to invasive growth and are not functional after the switch has been made. This possibility would explain why the *YMR181c* IRES did not increase expression from Tth-IRES-αHA, as the intron-integrated yeast cells assayed would have already finished their conversion to invasive growth.

<u>There is no relationship between intron circle formation and expression of Ppo.l1925 or Tth.L1925-encoded ORFs.</u>

Experiments deleting one of the IPS sites show that the full-length RNA is important for translation (Lin and Vogt 1998). The ribozyme region stabilizes the ORF and, the ribozyme could play a role in expression. It is possible that the message is a relatively rare species that requires the ribozyme for formation. Changes in the 5' end and 3'UTR sequence could be influencing formation of this hypothetical species and, thus influencing translation levels. One of the more appealing candidates is circular intron RNA, since this model would explain how sequences downstream of the ORF, such as the 3'UTR, influence translation of the ORF. After circularization,

the "downstream" sequences could recruit factors that would end up "upstream" of the ORF as well. Additionally, sequences near the 5' end of the intron and 3' end of the intron could influence the rate of circle formation and junction formed, which would also influence expression of the intron.

Both Ppo-α- and Tth-α-based introns form circular RNA species *in vivo*. Many of them form circles near the intron 5' end, although only two of the Tth-α introns, Tth-α and Tth-α-ran 3, appear to form full-length circles. Interestingly, all of the Tth-α-based introns form a circle between α nucleotides 288 and 289, while these similar circles were not seen in the Ppo-α-based introns. These circles may be related to the IGS circles formed by an ORF-less Tth.L1925 which, similarly, are not found in Ppo.L1925.

While all the introns tested form circular RNAs, it does not appear that circular intron RNAs are involved in the translation of the intron-encoded ORF. First, there is no obvious correlation between the circle junction and the amount of protein being translated from the intron RNA. Indeed, several of the high and low expressing introns shared the same junction. Secondly, many of the circle junctions are inside the α ORF sequence, removing the α start codon. There are no start codons in the 3' Tth.L1925 sequence that could act in place of the deleted start in intron circles, so it is impossible for these circles to be translated. Additionally, all of the Tth-α based introns form a circle that removes nearly all of the α sequence.

In conclusion, while intron sequences have an effect on the level of translation of an intron-encoded ORF, the mechanism by which they operate remains mysterious. In chapter IV, I found that there was no significant difference in RNA levels and processing between all of the Tth-α introns. In this chapter I found that it is unlikely that the intron RNAs are interacting with the invasive growth IRES pathway, since adding a known IRES sequence to an intron does not increase translation from the

intron in relation to adding the reverse complement of the selfsame IRES sequence. The formation of circular intron RNA also does not seem to play a role in translation of intron-encoded ORFS. While Tth-α and Ppo-α introns do form circular RNAs in yeast there is not a relationship between levels of expression of the intron-encoded ORF and the type of circle junction formed. Many of the highest expressing introns formed circles that disrupted the intron-encoded ORF.

The results of this chapter indicate that intron sequence must be changing expression levels through another mechanism. I will discuss a few of the possible mechanisms in the next chapter.

Chapter VI

Perspectives on Future Research

In this work, I have shown that a normally ORF-less group I intron, Tth.L1925, will express ORFs that are added to it and that the sequence of that intron can greatly influence the level of expression of the inserted ORF. However, the mechanism by which these introns are translated, and by which intron sequences influence translation, remains elusive. Based on my results, intron sequence does not influencing translation by altering the level of full-length intron RNA available for translation, by changing the amount of intron RNA processed at IPS1, IPS2, or IPS3, by interacting with the invasive growth IRES system, or by causing the intron RNA to form a highly-expressed circle. A number of questions exist that, if answered, could help illuminate how intron-encoded ORFs are translated.

Does the ribozyme have a role in expression of the intron-encoded ORF? The ribozyme structure certainly seems to stabilize the spliced intron RNA, as the intron RNAs that accumulate in the cell contain the ribozyme region (see Figures 4.4 and 4.15). This stabilizing effect, along with the large number of intron copies (~120) and the high activity of RNA Polymerase I, explains the large levels of intron RNA in the cell. Intron-derived RNA species in yeast amount to ~2% of the total RNA , an amount that is comparable to the sum total of all the mRNA in yeast (Warner 1971; Lin and Vogt 1998). Stabilization of the major intron RNA species, however, does not explain differences in expression, as all chimeric Tth.L1925 introns I have examined exhibit similar amounts of full-length and IPS-processed RNA, regardless of the expression level.

The activity of the ribozyme is required to splice the full-length intron RNA out of the pre-rRNA, but does it play a role in translation post-splicing? The enzymatic activity of the ribozyme could be required for translation of the intron-

encoded ORF. Although post-splicing processing and intron circularization do not appear to be involved in translation, a rare processed RNA could serve as the message for the intron-encoded protein. Such a hypothetical RNA species might be impossible to detect by northern blotting, however. Signal from the abundant full-length and IPS-processed RNA could drown out any signal.

The potential role of the reverse-splicing reaction in translation of intron-encoded ORFs has not been addressed. Since an RNA requires very little sequence identity to be a potential target for reverse-splicing, which requires base pairing with the some of the six nucleotides of the intron IGS (Roman and Woodson 1995; Roman, Rubin et al. 1999), it is likely that reverse splicing is occurring at some low rate in the yeast cells. Although the most abundant target RNA is the actual intron insertion site of the ribosomal large subunit RNA, it is likely that the intron is also able to reverse splice into other RNA species, including mRNA. By reverse splicing into an mRNA, the intron would gain both a cap and poly-A tail and, if it reverse-splices ahead of, or in frame with, the "native" start codon of the mRNA, the intron-encoded ORF could be translated just like a "normal" mRNA (Figure 6.1).

The reverse splicing model of translation is not ruled out by any of the experimental data. RNA species derived from reverse splicing do not appear on northern blots, but this could be because of their relative rarity compared to the overwhelming amount of full-length and processed intron RNA. It is possible that the "high-expression" sequences 5' and 3' of the intron-encoded ORF could increase translation, both by promoting the reverse splicing reaction and by allowing the intron to reverse splice at new sites.

If the intron is reverse splicing into mRNAs, it is likely that some of the sites are downstream of, and in frame with, an existing start codon. Proteins translated from these mRNA would be fusions, containing the N-terminal amino acids of the

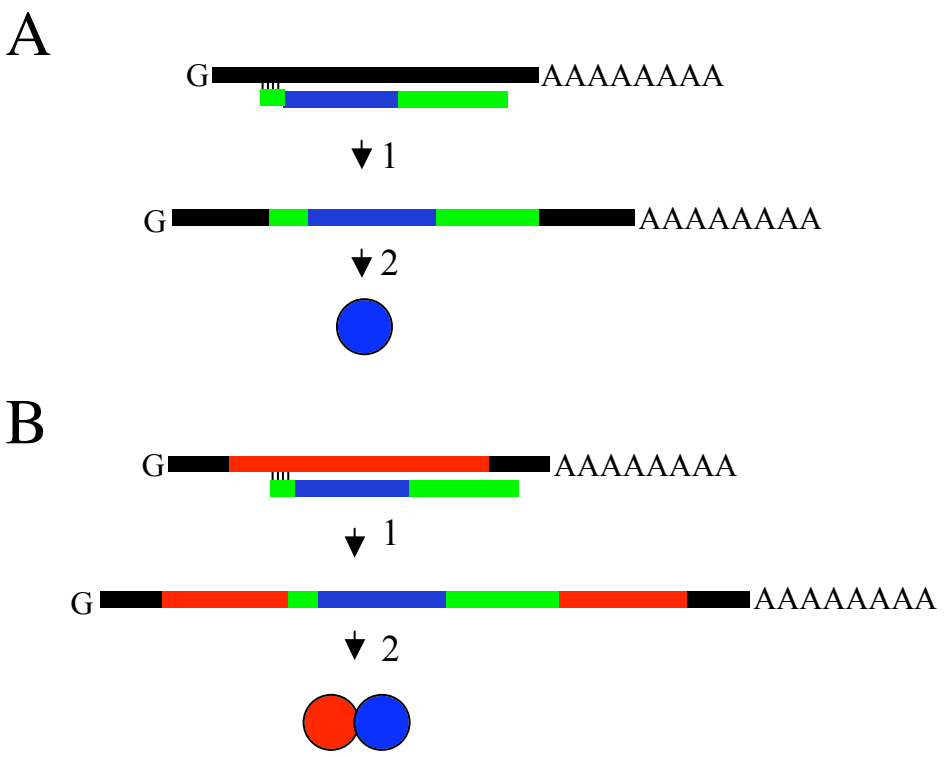


Figure 6.1: Hypothetical translation of intron encoded-ORFs through reverse splicing. A. 1. The intron (green) reverse splices at a site upstream of the mRNA's start codon. 2. The mRNA is translated and, since the start codon of the intron-encoded ORF(blue) is the first one encountered by the ribosome, the intron-encoded ORF is translated. B. 1. The intron reverse splices at a site downstream of, but in frame with, the ORF on the mRNA (red). 2. The mRNA is translated and produces a protein that is a fusion of the N-terminus of the mRNA-encoded protein and the entire intron-encoded protein.

mRNA-encoded protein fused to the entirety of the intron-encoded protein (Figure 6.1B). If this is happening with more than one mRNA, fusion proteins of varying lengths would be produced. Only a single protein species is seen in the western blot of protein from Tth-His3HA and Tth-5'P-His3HA (Figure 3.7 and 4.11). This result does not necessarily rule out the reverse-splicing model, though. Proteins corresponding to the full-length intron-encoded ORF can be translated from reverse-splicing at many sites, while each larger chimeric protein would be the result of reverse splicing at a single site (Figure 6.2). It is quite possible that the western blotting signal from chimeric protein(s) is drowned out by the signal of the full-length protein, explaining the single bands seen in Figure 3.7 and 4.11.

Models for translation of intron-encoded ORFs that do not require the enzymatic activity of the ribozyme are nebulous. Intron RNA could be forming a structure that interacts with cellular translation machinery. Sequences 5' and 3' of the ORF could promote the formation of this structure or discourage the formation of a competing, translationally inactive, structure. It is unclear what these structures might be or the particulars of the proteins the intron RNA would be interacting with, although perhaps chemical probing of high and low expressing intron RNAs would illuminate structural differences linked to expression.

An alternate model is that the intron RNA is interacting directly with the ribosomal RNA (Figure 6.3). The intron insertion site, which binds to the intron IGS, hangs over the active site of the ribosome (Figure 6.4). The first four of the six nucleotides bound by the IGS are unstructured in the 2000 LSU crystal structure and may be available to base-pair with the intron sequence (Ban, Nissen et al. 2000; Nissen, Hansen et al. 2000). Interaction of the intron RNA with these nucleotides could promote translation, perhaps by allowing a ribosome-associated intron RNA to displace the mRNA in a translation initiation complex. Intron sequences could alter

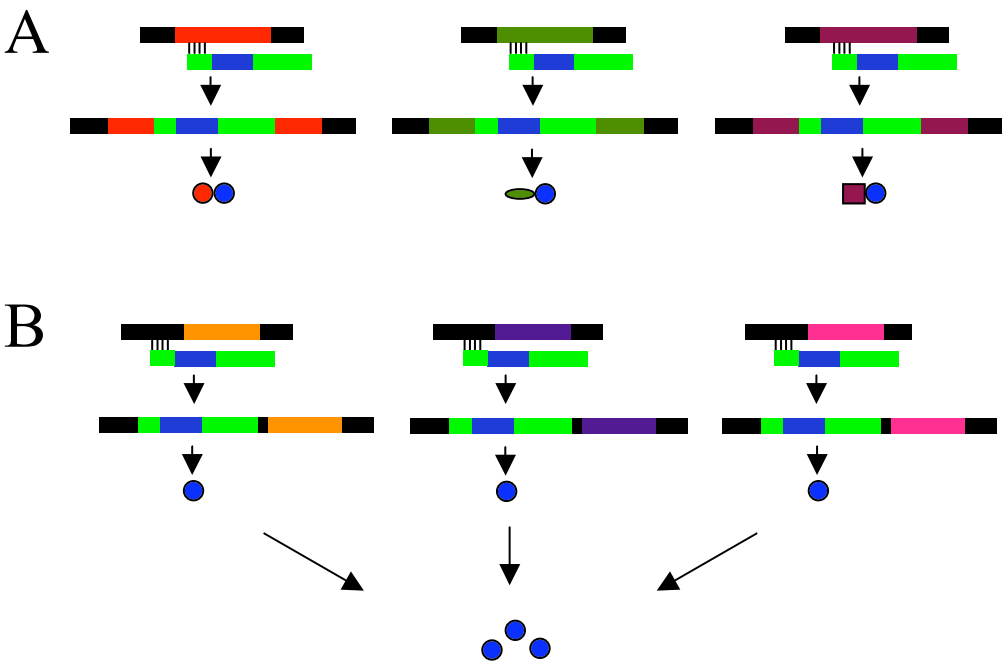
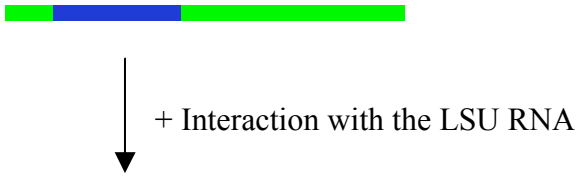


Figure 6.2: The full length intron-encoded protein could be the product of reverse splicing at many sites, while each fusion protein is the result of reverse splicing at a unique site. A. Each fusion protein produced through reverse splicing downstream of a start codon is unique. B. Introns that reverse splice upstream of a start codon all produce the full-length intron encoded protein.



Intron sequence could affect translation through:

•Changing the strength of interactions
between the intron RNA and LSU rRNA.

Figure 6.3: Model for intron-encoded ORF expression through interactions with the large ribosomal subunit RNA.

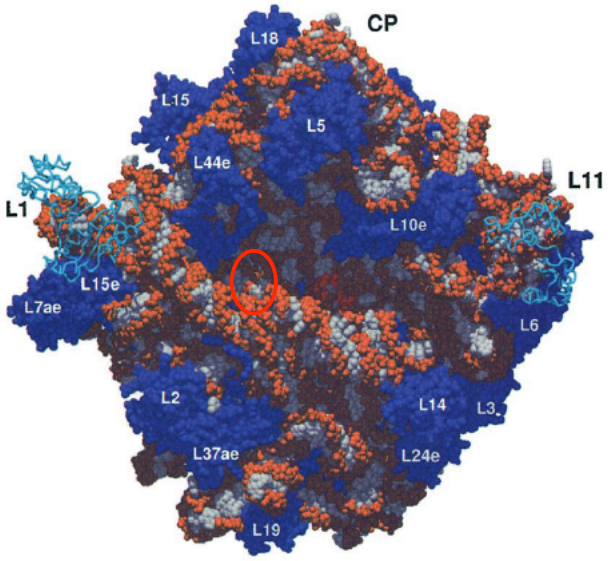


Figure 6.4: Space-filling model of the large ribosomal subunit. The circled region indicates the position of the intron insertion site. An inhibitor bound in the active site is indicated in red. The RNA bases are in grey and sugar-phosphate backbone in orange. Ribosomal proteins are in blue. Adapted from Nissen, Hansen et al., 2000.

expression levels by forming structures that promote or discourage ribosome-IGS binding.

Understanding how these introns-encoded ORFs are translated could be useful both to understand evolution and as a tool to produce RNA and express proteins *in vivo*. Since the cell contains so much intron RNA, if adjustments could be made in intron sequence to increase intron-based translation efficiency to that of Pol II mRNA, it could become a way to produce very large amounts of protein. There are major limitations of the system, however. Any ORF to be expressed could not disrupt intron splicing, a quality that has proven to be difficult to predict. Additionally, since Pol I translation is not inducible, the intron-encoded ORF would be constitutively expressed, which could cause difficulties in the case of proteins that are toxic at high levels. Perhaps modeling an intron after intron Dir.S956-2, which has an ORF oriented anti-sense to the intron sequence that is driven by an independent RNA Pol II promoter, would be a more useful tool (Figure 6.5). This arrangement would not be able to take advantage of the high activity of RNA Pol I, but would allow ~120 copies of the ORF of interest to be stably integrated, under the control of an inducible promoter, into the genome of the host cell.

A simpler use for these introns is as a vehicle to make large amounts of a small RNA irrespective of translation, for example, an RNA aptamer, which is an RNA sequences that forms a structure which binds to a specific molecule, *in vivo*. As with inserted ORFs, aptamer sequences would have to be tested to make sure they do not disrupt splicing, and, additionally, the intron sequences could bind to the aptamer and disrupt its interactions with the target molecule. Intron processing near the IGS could relieve intron-based disruption of the aptamer function, as the resulting 5' processed piece would include only minimal intron sequence.

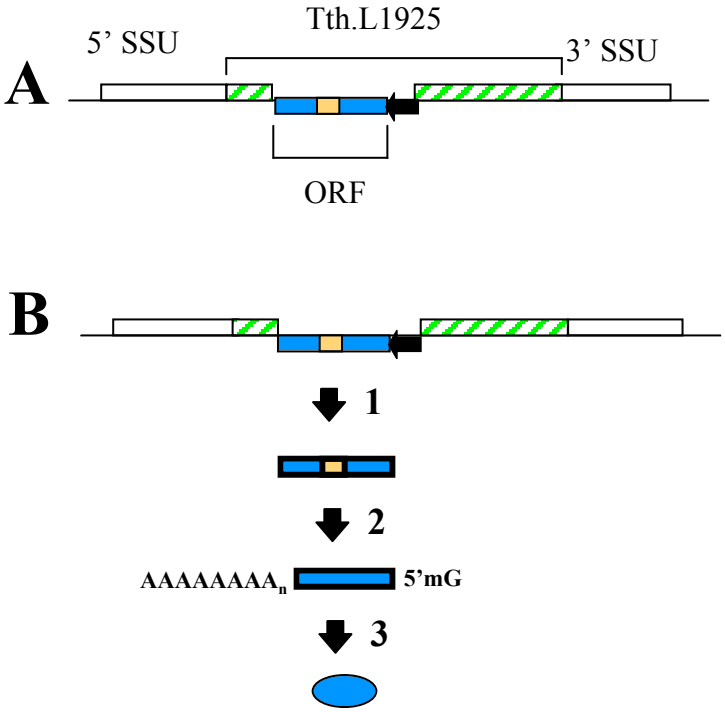


Figure 6.5: A Dir.S956-2-like model could be useful for increasing translation from chimeric Tth.L1925 introns. A. A schematic representation of a theoretical chimeric Tth.L1925. Note that the ORF (light blue) is oriented in the antisense direction in relation to Dir.S956-2. B. Expression of the intron-encoded ORF from a Dir.S956-2-like chimeric Tth.L1925. 1. The ORF DNA (thin lines) is transcribed by RNA Polymerase II into RNA (thick lines), driven by an inducible promoter (black arrow). This transcription is independent of the transcription of the SSU and intron. 2. An optional spliceosomal intron is removed by the spliceosomal complex, and a 5’ cap and poly-A tail is added to the I-DirII RNA. 3. The processed RNA is translated into the protein of interest (blue circle).

In conclusion, we do not understand many things about protein expression from nuclear group I introns. However, several tantalizing avenues for research have presented themselves. There are numerous reasons to pursue these avenues, as they could provide insight into novel pathways that lead from a DNA sequence to an expressed protein, provide greater understanding of the evolutionary history of mobile genetic elements, and provide a useful new tool to produce RNA and protein *in vivo*.

REFERENCES

Adams, P. L., M. R. Stahley, et al. (2004). "Crystal structure of a group I intron splicing intermediate." RNA **10**(12): 1867-87.

Adams, P. L., M. R. Stahley, et al. (2004). "Crystal structure of a self-splicing group I intron with both exons." Nature **430**(6995): 45-50.

Akins, R. A. and A. M. Lambowitz (1987). "A protein required for splicing group I introns in Neurospora mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof." Cell **50**(3): 331-45.

Altmann, M., N. Sonenberg, et al. (1989). "Translation in Saccharomyces cerevisiae: initiation factor 4E-dependent cell-free system." Mol Cell Biol **9**(10): 4467-72.

Amberg, D. C., D. J. Burke, et al. (2006) "Yeast Protein Extracts." Cold Spring Harbor Protocols **Volume**, DOI: 10.1101/pdb.prot4152

Ashe, M. P., S. K. De Long, et al. (2000). "Glucose depletion rapidly inhibits translation initiation in yeast." Mol Biol Cell **11**(3): 833-48.

Ausubel, F. M., R. Brent, et al. (1990). Current protocols in molecular biology. New York, NY, John Wiley and Sons, Inc.

Ban, N., P. Nissen, et al. (2000). "The complete atomic structure of the large ribosomal subunit at 2.4 A resolution." Science **289**(5481): 905-20.

Bassi, G. S., D. M. de Oliveira, et al. (2002). "Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA." Proc Natl Acad Sci U S A **99**(1): 128-33.

Been, M. D. and T. R. Cech (1986). "One binding site determines sequence specificity of Tetrahymena pre-rRNA self-splicing, trans-splicing, and RNA enzyme activity." Cell **47**(2): 207-16.

Been, M. D. and T. R. Cech (1987). "Selection of circularization sites in a group I IVS RNA requires multiple alignments of an internal template-like sequence." Cell **50**(6): 951-61.

Belfort, M. and P. S. Perlman (1995). "Mechanisms of intron mobility." J Biol Chem **270**(51): 30237-40.

Belfort, M. and R. J. Roberts (1997). "Homing endonucleases: keeping the house in order." Nucleic Acids Res **25**(17): 3379-88.

Bell-Pedersen, D., S. Quirk, et al. (1990). "Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications." Nucleic Acids Res **18**(13): 3763-70.

Bhattacarya, D., J. J. Cannone, et al. (2001). "Group I intron lateral transfer between red and brown algal ribosomal RNA." Curr Genet **40**(1): 82-90.

Birgisdottir, A. B. and S. Johansen (2005). "Site-specific reverse splicing of a HEG-containing group I intron in ribosomal RNA." Nucleic Acids Res **33**(6): 2042-51.

Bolduc, J. M., P. C. Spiegel, et al. (2003). "Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor." Genes Dev **17**(23): 2875-88.

Buchmueller, K. L., A. E. Webb, et al. (2000). "A collapsed non-native RNA folding state." Nat Struct Biol **7**(5): 362-6.

Burke, J. M. (1988). "Molecular genetics of group I introns: RNA structures and protein factors required for splicing--a review." Gene **73**(2): 273-94.

Burke, J. M. and U. L. RajBhandary (1982). "Intron within the large rRNA gene of N. crassa mitochondria: a long open reading frame and a consensus sequence possibly important in splicing." Cell **31**(3 Pt 2): 509-20.

Cannone, J. J., S. Subramanian, et al. (2002). "The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs." BMC Bioinformatics **3**: 2.

Caprara, M. G., V. Lehnert, et al. (1996). "A tyrosyl-tRNA synthetase recognizes a conserved tRNA-like structural motif in the group I intron catalytic core." Cell **87**(6): 1135-45.

Cech, T. R. (1986). "RNA as an enzyme." Sci Am **255**(5): 64-75.

Cech, T. R. (1988). "Conserved sequences and structures of group I introns: building an active site for RNA catalysis--a review." Gene **73**(2): 259-71.

Cech, T. R. (2000). "Structural biology. The ribosome is a ribozyme." Science **289**(5481): 878-9.

Cech, T. R., S. H. Damberger, et al. (1994). "Representation of the secondary and tertiary structure of group I introns." Nat Struct Biol **1**(5): 273-80.

Chatterjee, P., K. L. Brady, et al. (2003). "Functionally distinct nucleic acid binding sites for a group I intron encoded RNA maturase/DNA homing endonuclease." J Mol Biol **329**(2): 239-51.

Cho, Y., Y. L. Qiu, et al. (1998). "Explosive invasion of plant mitochondria by a group I intron." Proc Natl Acad Sci U S A **95**(24): 14244-9.

Christianson, T. W., R. S. Sikorski, et al. (1992). "Multifunctional yeast high-copy-number shuttle vectors." Gene **110**(1): 119-22.

Clodi, E., K. Semrad, et al. (1999). "Assaying RNA chaperone activity in vivo using a novel RNA folding trap." Embo J **18**(13): 3776-82.

Coetzee, T., D. Herschlag, et al. (1994). "Escherichia coli proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones." Genes Dev **8**(13): 1575-88.

Colleaux, L., L. d'Auriol, et al. (1986). "Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into E. coli as a specific double strand endonuclease." Cell **44**(4): 521-33.

Colleaux, L., L. D'Auriol, et al. (1988). "Recognition and cleavage site of the intron-encoded omega transposase." Proc Natl Acad Sci U S A **85**(16): 6022-6.

Cousineau, B., D. Smith, et al. (1998). "Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination." Cell **94**(4): 451-62.

Couture, S., A. D. Ellington, et al. (1990). "Mutational analysis of conserved nucleotides in a self-splicing group I intron." J Mol Biol **215**(3): 345-58.

Cullen, P. J. and G. F. Sprague, Jr. (2000). "Glucose depletion causes haploid invasive growth in yeast." Proc Natl Acad Sci U S A **97**(25): 13619-24.

Dalgaard, J. Z., A. J. Klar, et al. (1997). "Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family." Nucleic Acids Res **25**(22): 4626-38.

Davies, R. W., R. B. Waring, et al. (1982). "Making ends meet: a model for RNA splicing in fungal mitochondria." Nature **300**(5894): 719-24.

Decatur, W. A., C. Einvik, et al. (1995). "Two group I ribozymes with different functions in a nuclear rDNA intron." Embo J **14**(18): 4558-68.

Decatur, W. A., S. Johansen, et al. (2000). "Expression of the Naegleria intron endonuclease is dependent on a functional group I self-cleaving ribozyme." Rna **6**(4): 616-27.

Dickson, L., H. R. Huang, et al. (2001). "Retrotransposition of a yeast group II intron occurs by reverse splicing directly into ectopic DNA sites." Proc Natl Acad Sci U S A **98**(23): 13207-12.

Doudna, J. A., B. P. Cormack, et al. (1989). "RNA structure, not sequence, determines the 5' splice-site specificity of a group I intron." Proc Natl Acad Sci U S A **86**(19): 7402-6.

Dujon, B., M. Belfort, et al. (1989). "Mobile introns: definition of terms and recommended nomenclature." Gene **82**(1): 115-8.

Einvik, C., W. A. Decatur, et al. (1997). "Naegleria nucleolar introns contain two group I ribozymes with different functions in RNA splicing and processing." Rna **3**(7): 710-20.

Eskes, R., L. Liu, et al. (2000). "Multiple homing pathways used by yeast mitochondrial group II introns." Mol Cell Biol **20**(22): 8432-46.

Eskes, R., J. Yang, et al. (1997). "Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing." Cell **88**(6): 865-74.

Esser, K., E. Pratje, et al. (1996). "SOM 1, a small new gene required for mitochondrial inner membrane peptidase function in Saccharomyces cerevisiae." Mol Gen Genet **252**(4): 437-45.

Field, J., J. Nikawa, et al. (1988). "Purification of a RAS-responsive adenylyl cyclase complex from Saccharomyces cerevisiae by use of an epitope addition method." Mol Cell Biol **8**(5): 2159-65.

Fleischer, S. and I. Grummt (1983). "Expression of an mRNA coding gene under the control of an RNA polymerase I promoter." Embo J **2**(12): 2319-22.

Flick, K. E., M. S. Jurica, et al. (1998). "DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI." Nature **394**(6688): 96-101.

Gallie, D. R. (1991). "The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency." Genes Dev **5**(11): 2108-16.

Gampel, A., M. Nishikimi, et al. (1989). "CBP2 protein promotes in vitro excision of a yeast mitochondrial group I intron." Mol Cell Biol **9**(12): 5424-33.

Gilbert, W. V., K. Zhou, et al. (2007). "Cap-independent translation is required for starvation-induced differentiation in yeast." Science **317**(5842): 1224-7.

Gimeno, C. J., P. O. Ljungdahl, et al. (1992). "Unipolar cell divisions in the yeast S. cerevisiae lead to filamentous growth: regulation by starvation and RAS." Cell **68**(6): 1077-90.

Golden, B. L., A. R. Gooding, et al. (1998). "A preorganized active site in the crystal structure of the Tetrahymena ribozyme." Science **282**(5387): 259-64.

Golden, B. L., H. Kim, et al. (2005). "Crystal structure of a phage Twort group I ribozyme-product complex." Nat Struct Mol Biol **12**(1): 82-9.

Green, R. and J. W. Szostak (1992). "Selection of a ribozyme that functions as a superior template in a self-copying reaction." Science **258**(5090): 1910-5.

Grummt, I. and J. A. Skinner (1985). "Efficient transcription of a protein-coding gene from the RNA polymerase I promoter in transfected cells." Proc Natl Acad Sci U S A **82**(3): 722-6.

Guerrier-Takada, C., K. Gardiner, et al. (1983). "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme." Cell **35**(3 Pt 2): 849-57.

Gunnery, S. and M. B. Mathews (1995). "Functional mRNA can be generated by RNA polymerase III." Mol Cell Biol **15**(7): 3597-607.

Guo, F., A. R. Gooding, et al. (2004). "Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site." Mol Cell **16**(3): 351-62.

Hanna, R. and J. A. Doudna (2000). "Metal ions in ribozyme folding and catalysis." Curr Opin Chem Biol **4**(2): 166-70.

Haugen, P., J. F. De Jonckheere, et al. (2002). "Characterization of the self-splicing products of two complex Naegleria LSU rDNA group I introns containing homing endonuclease genes." Eur J Biochem **269**(6): 1641-9.

Haugen, P., O. G. Wikmark, et al. (2005). "The recent transfer of a homing endonuclease gene." Nucleic Acids Res **33**(8): 2734-41.

He, F., X. Li, et al. (2003). "Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast." Mol Cell **12**(6): 1439-52.

Herbert, C. J., M. Labouesse, et al. (1988). "The NAM2 proteins from S. cerevisiae and S. douglasii are mitochondrial leucyl-tRNA synthetases, and are involved in mRNA splicing." Embo J **7**(2): 473-83.

Herschlag, D. (1995). "RNA chaperones and the RNA folding problem." J Biol Chem **270**(36): 20871-4.

Hetzer, M., G. Wurzer, et al. (1997). "Trans-activation of group II intron splicing by nuclear U5 snRNA." Nature **386**(6623): 417-20.

Hill, J., P. McGraw, et al. (1985). "A mutation in yeast mitochondrial DNA results in a precise excision of the terminal intron of the cytochrome b gene." J Biol Chem **260**(6): 3235-8.

Holst-Jensen, A., M. Vaage, et al. (1999). "Structural characteristics and possible horizontal transfer of group I introns between closely related plant pathogenic fungi." Mol Biol Evol **16**(1): 114-26.

Imataka, H., A. Gradi, et al. (1998). "A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation." Embo J **17**(24): 7480-9.

Inoue, T., F. X. Sullivan, et al. (1986). "New reactions of the ribosomal RNA precursor of Tetrahymena and the mechanism of self-splicing." J Mol Biol **189**(1): 143-65.

Jackson, S. A., S. Koduvayur, et al. (2006). "Self-splicing of a group I intron reveals partitioning of native and misfolded RNA populations in yeast." RNA **12**(12): 2149-59.

Johansen, S., M. Elde, et al. (1997). "In vivo mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete Didymium iridis." Mol Microbiol **24**(4): 737-45.

Johansen, S. and P. Haugen (2001). "A new nomenclature of group I introns in ribosomal DNA." Rna **7**(7): 935-6.

Johansen, S. and V. M. Vogt (1994). "An intron in the nuclear ribosomal DNA of Didymium iridis codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing." Cell **76**(4): 725-34.

Johansen, S. D., P. Haugen, et al. (2007). "Expression of protein-coding genes embedded in ribosomal DNA." Biol Chem **388**(7): 679-86.

Johansen, S. D., A. Vader, et al. (2006). "In vivo expression of a group I intron HEG from the antisense strand of Didymium ribosomal DNA." RNA Biol **3**(4): 157-62.

Kessler, S. H. and A. B. Sachs (1998). "RNA recognition motif 2 of yeast Pab1p is required for its functional interaction with eukaryotic translation initiation factor 4G." Mol Cell Biol **18**(1): 51-7.

Khaitovich, P., A. S. Mankin, et al. (1999). "Characterization of functionally active subribosomal particles from Thermus aquaticus." Proc Natl Acad Sci U S A **96**(1): 85-90.

Khaitovich, P., T. Tenson, et al. (1999). "Peptidyl transferase activity catalyzed by protein-free 23S ribosomal RNA remains elusive." Rna **5**(5): 605-8.

Koduvayur, S. P. and S. A. Woodson (2004). "Intracellular folding of the Tetrahymena group I intron depends on exon sequence and promoter choice." RNA **10**(10): 1526-32.

Komar, A. A., T. Lesnik, et al. (2003). "Internal initiation drives the synthesis of Ure2 protein lacking the prion domain and affects [URE3] propagation in yeast cells." EMBO J **22**(5): 1199-209.

Kreike, J., M. Schulze, et al. (1987). "A yeast nuclear gene, MRS1, involved in mitochondrial RNA splicing: nucleotide sequence and mutational analysis of two overlapping open reading frames on opposite strands." Embo J **6**(7): 2123-9.

Kreike, J., M. Schulze, et al. (1986). "Cloning of a nuclear gene MRS1 involved in the excision of a single group I intron (bI3) from the mitochondrial COB transcript in S. cerevisiae." Curr Genet **11**(3): 185-91.

Kruger, K., P. J. Grabowski, et al. (1982). "Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena." Cell **31**(1): 147-57.

Labouesse, M., C. J. Herbert, et al. (1987). "Three suppressor mutations which cure a mitochondrial RNA maturase deficiency occur at the same codon in the open reading frame of the nuclear NAM2 gene." Embo J **6**(3): 713-21.

Lambowitz, A. M. (1989). "Infectious introns." Cell **56**(3): 323-6.

Lambowitz, A. M. and M. Belfort (1993). "Introns as mobile genetic elements." Annu Rev Biochem **62**: 587-622.

Lee, N., Y. Bessho, et al. (2000). "Ribozyme-catalyzed tRNA aminoacylation." Nat Struct Biol **7**(1): 28-33.

Lin, J. and V. M. Vogt (1998). "I-PpoI, the endonuclease encoded by the group I intron PpLSU3, is expressed from an RNA polymerase I transcript." Mol Cell Biol **18**(10): 5809-17.

Lin, J. and V. M. Vogt (2000). "Functional alpha-fragment of beta-galactosidase can be expressed from the mobile group I intron PpLSU3 embedded in yeast pre-ribosomal RNA derived from the chromosomal rDNA locus." Nucleic Acids Res **28**(6): 1428-38.

Liu, X. Q. (2000). "Protein-splicing intein: Genetic mobility, origin, and evolution." Annu Rev Genet **34**: 61-76.

Lo, H. J., H. K. Huang, et al. (1998). "RNA polymerase I-promoted HIS4 expression yields uncapped, polyadenylated mRNA that is unstable and inefficiently translated in Saccharomyces cerevisiae." Mol Cell Biol **18**(2): 665-75.

Lohse, P. A. and J. W. Szostak (1996). "Ribozyme-catalysed amino-acid transfer reactions." Nature **381**(6581): 442-4.

Lopata, M. A., D. W. Cleveland, et al. (1986). "RNA polymerase specificity of mRNA production and enhancer action." Proc Natl Acad Sci U S A **83**(18): 6677-81.

McGraw, P. and A. Tzagoloff (1983). "Assembly of the mitochondrial membrane system. Characterization of a yeast nuclear gene involved in the processing of the cytochrome b pre-mRNA." J Biol Chem **258**(15): 9459-68.

Michel, F., A. Jacquier, et al. (1982). "Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure." Biochimie **64**(10): 867-81.

Michel, F. and E. Westhof (1990). "Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis." J Mol Biol **216**(3): 585-610.

Mohr, G., M. G. Caprara, et al. (1994). "A tyrosyl-tRNA synthetase can function similarly to an RNA structure in the Tetrahymena ribozyme." Nature **370**(6485): 147-50.

Mohr, G., A. Zhang, et al. (1992). "The neurospora CYT-18 protein suppresses defects in the phage T4 td intron by stabilizing the catalytically active structure of the intron core." Cell **69**(3): 483-94.

Moore, M. J. and P. A. Sharp (1993). "Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing." Nature **365**(6444): 364-8.

Moran, J. V., S. Zimmerly, et al. (1995). "Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements." Mol Cell Biol **15**(5): 2828-38.

Morino, S., H. Hazama, et al. (1996). "Analysis of the mRNA cap-binding ability of human eukaryotic initiation factor-4E by use of recombinant wild-type and mutant forms." Eur J Biochem **239**(3): 597-601.

Mota, E. M. and R. A. Collins (1988). "Independent evolution of structural and coding regions in a Neurospora mitochondrial intron." Nature **332**(6165): 654-6.

Mueller, J. E., Bryk, M., Loizos, N. and Belfort, M. (1993). Nucleases. S. M. Linn, Lloyd, R.S., and Roberts, R.J. Cold Spring Harbor, Cold Spring Harbor Press**:** 111-143.

Muscarella, D. E. and V. M. Vogt (1989). "A mobile group I intron in the nuclear rDNA of Physarum polycephalum." Cell **56**(3): 443-54.

Muscarella, D. E. and V. M. Vogt (1993). "A mobile group I intron from Physarum polycephalum can insert itself and induce point mutations in the nuclear ribosomal DNA of saccharomyces cerevisiae." Mol Cell Biol **13**(2): 1023-33.

Nam, H.-S. (1999). Expression of SOM1 from a Group I intron in the rDNA of Saccharomyces cerevisiae. Section of Biochemistry, Cell and Molecular Biology. Ithaca, Cornell University. **BA Honors Thesis**.

Nielsen, H. and J. Engberg (1985). "Sequence comparison of the rDNA introns from six different species of Tetrahymena." Nucleic Acids Res **13**(20): 7445-55.

Nielsen, H., T. Fiskaa, et al. (2003). "The ability to form full-length intron RNA circles is a general property of nuclear group I introns." RNA **9**(12): 1464-75.

Nielsen, H., E. Westhof, et al. (2005). "An mRNA is capped by a 2', 5' lariat catalyzed by a group I-like ribozyme." Science **309**(5740): 1584-7.

Nissen, P., J. Hansen, et al. (2000). "The structural basis of ribosome activity in peptide bond synthesis." Science **289**(5481): 920-30.

Padgett, R. A., M. Podar, et al. (1994). "The stereochemical course of group II intron self-splicing." Science **266**(5191): 1685-8.

Palecek, S. P., A. S. Parikh, et al. (2002). "Depression of Saccharomyces cerevisiae invasive growth on non-glucose carbon sources requires the Snf1 kinase." Mol Microbiol **45**(2): 453-69.

Paukstelis, P. J., J. H. Chen, et al. (2008). "Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA." Nature **451**(7174): 94-7.

Perler, F. B., G. J. Olsen, et al. (1997). "Compilation and analysis of intein sequences." Nucleic Acids Res **25**(6): 1087-93.

Pyle, A. M. (1993). "Ribozymes: a distinct class of metalloenzymes." Science **261**(5122): 709-14.

Raychaudhuri, S., V. Fontanes, et al. (2006). "Zuotin, a DnaJ molecular chaperone, stimulates cap-independent translation in yeast." Biochem Biophys Res Commun **350**(3): 788-95.

Reineke, L. C., A. A. Komar, et al. (2008). "A small stem loop element directs internal initiation of the URE2 internal ribosome entry site in Saccharomyces cerevisiae." J Biol Chem **283**(27): 19011-25.

Rho, S. B. and S. A. Martinis (2000). "The bI4 group I intron binds directly to both its protein splicing partners, a tRNA synthetase and maturase, to facilitate RNA splicing activity." Rna **6**(12): 1882-94.

Roberts, R. J., and Halford, S.E. (1993). Nucleases. S. M. Linn, Lloyd, R.S., and Roberts, R.J. Cold Spring Harbor, Cold Spring Harbor Press**:** 35-88.

Roberts, R. J., M. Belfort, et al. (2003). "A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes." Nucleic Acids Res **31**(7): 1805-12.

Roberts, R. J. and D. Macelis (1997). "REBASE-restriction enzymes and methylases." Nucleic Acids Res **25**(1): 248-62.

Rocheleau, G. A. and S. A. Woodson (1995). "Enhanced self-splicing of Physarum polycephalum intron 3 by a second group I intron." Rna **1**(2): 183-93.

Roman, J., M. N. Rubin, et al. (1999). "Sequence specificity of in vivo reverse splicing of the Tetrahymena group I intron." RNA **5**(1): 1-13.

Roman, J. and S. A. Woodson (1995). "Reverse splicing of the Tetrahymena IVS: evidence for multiple reaction sites in the 23S rRNA." Rna **1**(5): 478-90.

Ruoff, B., S. Johansen, et al. (1992). "Characterization of the self-splicing products of a mobile intron from the nuclear rDNA of Physarum polycephalum." Nucleic Acids Res **20**(22): 5899-906.

Sachs, A. B., R. W. Davis, et al. (1987). "A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability." Mol Cell Biol **7**(9): 3268-76.

Sachs, A. B. and G. Varani (2000). "Eukaryotic translation initiation: there are (at least) two sides to every story." Nat Struct Biol **7**(5): 356-61.

Saguez, C., G. Lecellier, et al. (2000). "Intronic GIY-YIG endonuclease gene in the mitochondrial genome of Podospora curvicolla: evidence for mobility." Nucleic Acids Res **28**(6): 1299-306.

Sambrook, J. E., F. Fritsch, et al. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Schmitt, M. E., T. A. Brown, et al. (1990). "A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae." Nucleic Acids Res **18**(10): 3091-2.

Seino, A., Y. Yanagida, et al. (2005). "Translational control by internal ribosome entry site in Saccharomyces cerevisiae." Biochim Biophys Acta **1681**(2-3): 166-74.

Sellem, C. H. and L. Belcour (1997). "Intron open reading frames as mobile elements and evolution of a group I intron." Mol Biol Evol **14**(5): 518-26.

Semrad, K. and R. Schroeder (1998). "A ribosomal function is necessary for efficient splicing of the T4 phage thymidylate synthase intron in vivo." Genes Dev **12**(9): 1327-37.

Sharp, P. A. (1985). "On the origin of RNA splicing and introns." Cell **42**(2): 397-400.

Sharp, P. A. (1991). ""Five easy pieces"." Science **254**(5032): 663.

Shaw, L. C. and A. S. Lewin (1997). "The Cbp2 protein stimulates the splicing of the omega intron of yeast mitochondria." Nucleic Acids Res **25**(8): 1597-604.

Sherman, F., G. R. Fink, et al. (1986). Laboratory course manual for methods in yeast genetics. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Shub, D. A., J. M. Gott, et al. (1988). "Structural conservation among three homologous introns of bacteriophage T4 and the group I introns of eukaryotes." Proc Natl Acad Sci U S A **85**(4): 1151-5.

Shukla, G. C. and R. A. Padgett (2002). "A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome." Mol Cell **9**(5): 1145-50.

Smale, S. T. and R. Tjian (1985). "Transcription of herpes simplex virus tk sequences under the control of wild-type and mutant human RNA polymerase I promoters." Mol Cell Biol **5**(2): 352-62.

Sogin, M. L., A. Ingold, et al. (1986). "Phylogenetic evidence for the acquisition of ribosomal RNA introns subsequent to the divergence of some of the major Tetrahymena groups." Embo J **5**(13): 3625-30.

Steers, E., Jr., P. Cuatrecasas, et al. (1971). "The purification of beta-galactosidase from Escherichia coli by affinity chromatography." J Biol Chem **246**(1): 196-200.

Suzuki, C., Y. Hori, et al. (2003). "Screening and characterization of transposon-insertion mutants in a pseudohyphal strain of Saccharomyces cerevisiae." Yeast **20**(5): 407-15.

Tarun, S. Z., Jr. and A. B. Sachs (1997). "Binding of eukaryotic translation initiation factor 4E (eIF4E) to eIF4G represses translation of uncapped mRNA." Mol Cell Biol **17**(12): 6876-86.

Tarun, S. Z., Jr., S. E. Wells, et al. (1997). "Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation." Proc Natl Acad Sci U S A **94**(17): 9046-51.

Thompson, A. J. and D. L. Herrin (1994). "A chloroplast group I intron undergoes the first step of reverse splicing into host cytoplasmic 5.8 S rRNA. Implications for intron-mediated RNA recombination, intron transposition and 5.8 S rRNA structure." J Mol Biol **236**(2): 455-68.

Thompson, S. R., K. D. Gulyas, et al. (2001). "Internal initiation in Saccharomyces cerevisiae mediated by an initiator tRNA/eIF2-independent internal ribosome entry site element." Proc Natl Acad Sci U S A **98**(23): 12972-7.

Tian, G. L., G. Y. Li, et al. (1998). "The novel function of the Saccharomyces cerevisiae CBP2 gene as a splicing factor essential to excision of the Saccharomyces douglasii LSU intron in vivo." Mol Gen Genet **258**(1-2): 60-8.

Vader, A., S. Johansen, et al. (2002). "The group I-like ribozyme DiGIR1 mediates alternative processing of pre-rRNA transcripts in Didymium iridis." Eur J Biochem **269**(23): 5804-12.

Vader, A., H. Nielsen, et al. (1999). "In vivo expression of the nucleolar group I intron-encoded I-dirI homing endonuclease involves the removal of a spliceosomal intron." Embo J **18**(4): 1003-13.

Vallett, F., E. Merge, et al. (1988). "Construction of mutant and chimeric genes using the polymerase chain reaction." Nucleic Acids Res **17**: 723-726.

Wallweber, G. J., S. Mohr, et al. (1997). "Characterization of Neurospora mitochondrial group I introns reveals different CYT-18 dependent and independent splicing strategies and an alternative 3' splice site for an intron ORF." Rna **3**(2): 114-31.

Wardleworth, B. N., M. Kvaratskhelia, et al. (2000). "Site-directed mutagenesis of the yeast resolving enzyme Cce1 reveals catalytic residues and relationship with the intron-splicing factor Mrs1." J Biol Chem **275**(31): 23725-8.

Warner, J. R. (1971). "The assembly of ribosomes in yeast." J Biol Chem **246**(2): 447-54.

Webb, A. E., M. A. Rose, et al. (2001). "Protein-dependent transition states for ribonucleoprotein assembly." J Mol Biol **309**(5): 1087-100.

Webb, A. E. and K. M. Weeks (2001). "A collapsed state functions to self-chaperone RNA folding into a native ribonucleoprotein complex." Nat Struct Biol **8**(2): 135-40.

Weeks, K. M. and T. R. Cech (1995). "Efficient protein-facilitated splicing of the yeast mitochondrial bI5 intron." Biochemistry **34**(23): 7728-38.

Weeks, K. M. and T. R. Cech (1996). "Assembly of a ribonucleoprotein catalyst by tertiary structure capture." Science **271**(5247): 345-8.

Weinstein, L. B., B. C. Jones, et al. (1997). "A second catalytic metal ion in group I ribozyme." Nature **388**(6644): 805-8.

Woodson, S. A. and T. R. Cech (1989). "Reverse self-splicing of the tetrahymena group I intron: implication for the directionality of splicing and for intron transposition." Cell **57**(2): 335-45.

Woodson, S. A. and T. R. Cech (1991). "Alternative secondary structures in the 5' exon affect both forward and reverse self-splicing of the Tetrahymena intervening sequence RNA." Biochemistry **30**(8): 2042-50.

Zaug, A. J., P. J. Grabowski, et al. (1983). "Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage-ligation reaction." Nature **301**(5901): 578-83.

Zhang, A., V. Derbyshire, et al. (1995). "Escherichia coli protein StpA stimulates self-splicing by promoting RNA assembly in vitro." Rna **1**(8): 783-93.

Zhang, F., E. S. Ramsay, et al. (1995). "In vivo facilitation of Tetrahymena group I intron splicing in Escherichia coli pre-ribosomal RNA." Rna **1**(3): 284-92.

Zhou, W., G. M. Edelman, et al. (2001). "Transcript leader regions of two Saccharomyces cerevisiae mRNAs contain internal ribosome entry sites that function in living cells." Proc Natl Acad Sci U S A **98**(4): 1531-6.

Zimmerly, S., H. Guo, et al. (1995). "A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility." Cell **83**(4): 529-38.

Zimmerly, S., H. Guo, et al. (1995). "Group II intron mobility occurs by target DNA-primed reverse transcription." Cell **82**(4): 545-54.